

DETERMINANTS OF VESICULAR STOMATITIS VIRUS AND CORONAVIRUS  
FUSION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Shoshannah Lee Roth

August 2009

© 2009 Shoshannah Lee Roth

# DETERMINANTS OF VESICULAR STOMATITIS VIRUS AND CORONAVIRUS FUSION

Shoshannah Lee Roth, Ph. D.

Cornell University 2009

Vesicular stomatitis virus (VSV) is the prototypic and best-studied member of the *Rhabdoviridae*. Despite this, there remain many uncertainties regarding the exact process by which VSV enters host cells. One proposed mechanism involves virus internalization into endosomes, followed by a two-step fusion reaction; initial fusion with internal vesicles of multivesicular endosomes followed by subsequent back-fusion of these vesicles into the cytoplasm. However, the biological rationale for this pathway remains obscure. One feature of the internal vesicles of late endosomes is that they uniquely contain the lipid lysobisphosphatidic acid (LBPA). Using a FRET-based *in vitro* lipid mixing assay, we show that the presence of LBPA significantly increases the rate of VSV G-mediated membrane fusion. The increased rate of lipid mixing was selective for VSV and was not evident when other viruses, such as influenza, were examined in the same assay. These data provide a biological rationale for a two-step fusion reaction during VSV entry, and suggests that LBPA may preferentially affect the ability of VSV G (a class III viral fusion protein) to mediate lipid mixing during membrane fusion.

The *Coronaviridae* is a virus family with significant potential to negatively affect the human population. An improvement in our understanding of the entry mechanism of these viruses could assist in the development of therapeutics. Here we characterize a proposed fusion peptide in several members of the coronavirus family.

Our results show that this peptide is capable of promoting membrane fusion in a liposome based assay. When the sequence of the hypothesized fusion peptide is manipulated, its ability to promote lipid mixing is abolished. Additionally, we outline significant differences between the fusion peptides of several different coronavirus family members and postulate the effect these differences have on virus fusion. Our results support the hypothesis that this region is in fact the fusion peptide for the coronavirus family.

## BIOGRAPHICAL SKETCH

Shoshannah Lee Roth was born and raised in Iowa City, Iowa. She attended Iowa City high school where she played violin in the orchestra program and was an active member of several clubs including model United Nations, the science club and the debate team. While in high school she spent her summers doing research in the laboratory of Dr. Peter A. Rubenstein in the Department of Biochemistry at the University of Iowa making interesting mutants of yeast actin. In 1998, she graduated from high school and enrolled in the University of Iowa after receiving Presidential and Opportunity at Iowa scholarships. During her undergraduate career she participated in research in the Department of Chemistry in the laboratory of Dr. James B. Gloer determining the identity of antifungal metabolites of *Acremonium zeae*, a common fungus of corn, in collaboration with the USDA. She also participated in two summer research opportunities at in the laboratory of Dr. Marion Stankovich in the Department of Chemistry at the University of Minnesota-Twin Cities, and in the laboratory of Dr. Robert S. Coleman in the Department of Chemistry at The Ohio State University. In 2002, she graduated with highest honors with a B. S. in chemistry and a B. A. in psychology and began her graduate career at Cornell University in the Department of Chemistry and Chemical Biology. She began her research in the laboratory of Dr. D. Tyler McQuade developing directed evolution methods to modify enzymes to catalyze novel reactions and earned her M. S. in August of 2004. In August of 2007 she transferred to the Department of Microbiology and Immunology and began research in the laboratory of Dr. Garry R. Whittaker.

Upon completion of her Ph. D. work, Shannah intends to pursue post-doctoral research in the laboratory of Dr. Fredric Bushman at the University of Pennsylvania Medical School. Her work will examine the mechanism of DNA integration by the HIV integrase protein.

## ACKNOWLEDGMENTS

In particular, I would like to thank Gary Whittaker for his support and help in the last years of my graduate career. Gary accepted me in to his laboratory group when I knew little about virology and less about entry and fusion. He believed in me and allowed me to attempt research I didn't fully understand. Most of all he was patient while I endeavored to learn an entirely different field and supportive when things didn't go smoothly. I have no doubt that without him I could not have continued on the road to completing my graduate studies.

I would also like to thank my special committee members Dr. James Casey, Dr. Richard Cerione, and Dr. Ruth Collins for their helpful advice on my research. They have been understanding of my unique situation and extremely helpful and patient.

I would also like to thank the entire past and present Whittaker group: Ikenna Madu, Sandrine Belouzard, Xiangjie Sun, Yueting Zhang, Michele Bialecki, Victor, Chu, Victor Tse, Andrew Regan and Beth Licitra for helping me learn techniques and think through experiments as well as keeping my spirits up when research was getting me down. Additionally, I'd like to thank A. Damon Ferguson, Nadia Chapman and Dave Ousterout for technical support. I'd also like to thank the faculty and staff of the Department of Microbiology and Immunology and the College of Veterinary Medicine including Janna Lamey, Casey Isham, Mary Linton, and Walter Iddings for all of their help with the transition to the department and the non-science aspects of being a graduate student.

Finally, I'd like to acknowledge my parents and sisters for their love and support. Especially my mom for being just a phone call away when I hit a rough patch, and my sister Aleeza for our conversations about science and graduate school. I

am also most grateful for the love and support of my fiancé Jason Riggs. I believe that our relationship is stronger for us having gone through the challenges of graduate school together and emerging triumphant on the other side.

*“There is a single light of science and to brighten it anywhere is to brighten it everywhere”—Issac Asimov*

## TABLE OF CONTENTS

|  |      |
|--|------|
| Biographical sketch  | iii  |
| Acknowledgements   | iv   |
| Table of contents  | vi   |
| List of figures  | vii  |
| List of tables   | viii |
| Chapter 1: Introduction  | 1    |
| Chapter 2: Promotion of viral membrane fusion by the endosome-specific<br>phospholipid lysobisphosphatidic acid (LBPA) | 26   |
| Chapter 3: Lipid mixing promoted by novel coronavirus fusion peptides  | 43   |
| Chapter 4: Summary and conclusions   | 73   |
| Appendix 1: The effect of N-glycosylation of human cells on<br>influenza virus infection                               | 76   |
| Appendix 2: Directed evolution of Stoffel Fragment DNA polymerase  | 89   |



## LIST OF FIGURES

### Chapter 1

|            |  |    |
|------------|--|----|
| Figure 1.1 | Entry mechanisms used by viruses         | 3  |
| Figure 1.2 | The endocytic pathway                    | 6  |
| Figure 1.3 | Membrane fusion in the endocytic pathway | 8  |
| Figure 1.4 | Vesicular stomatitis virus structure     | 9  |
| Figure 1.5 | Proposed VSV entry pathways              | 12 |
| Figure 1.6 | Coronavirus structure                    | 13 |
| Figure 1.7 | Coronavirus S protein                    | 14 |
| Figure 1.8 | Influenza A structure                    | 16 |

### Chapter 2

|            |   |    |
|------------|---|----|
| Figure 2.1 | Lipid mixing of liposomes by VSV  | 28 |
| Figure 2.2 | Rate of lipid mixing promoted by VSV from pH 5 to 6.5                       | 29 |
| Figure 2.3 | Rate of lipid mixing promoted by VSV with changes in<br>LBPA                | 30 |
| Figure 2.4 | Comparison of VSV and influenza virus lipid mixing                          | 31 |
| Figure 2.5 | Effect of ionic lipids on VSV promoted lipid mixing                         | 33 |
| Figure 2.6 | Comparison of the VSV-Ind and VSV-NJ lipid mixing                           | 35 |
| Figure 2.7 | Comparison of VSV lipid mixing at pH 5 and 5.5 for<br>both serotypes of VSV | 36 |

### Chapter 3

|            |  |    |
|------------|--|----|
| Figure 3.1 | Fusion assays involving liposomes                      | 47 |
| Figure 3.2 | Alignment of coronavirus S proteins                    | 51 |
| Figure 3.3 | Lipid mixing by the IBV fusion peptide                 | 56 |
| Figure 3.4 | Lipid mixing by the SARS fusion peptide                | 57 |
| Figure 3.5 | Lipid mixing by the SARS fusion peptide from pH 5 to 7 | 58 |

|             |  |     |
|-------------|--|-----|
| Figure 3.6  | Lipid mixing by the modified SARS fusion peptide, pH 7                   | 59  |
| Figure 3.7  | Lipid mixing by the modified SARS fusion peptide, pH 5                   | 60  |
| Figure 3.8  | Lipid mixing by FCoV fusion peptides, pH 5                               | 61  |
| Figure 3.9  | Lipid mixing by FCoV fusion peptides, pH 7                               | 62  |
| Appendix 1  |  |     |
| Figure A1.1 | Overview of siRNA pathway  | 80  |
| Figure A1.2 | Expression of MGAT1 after first siRNA transfection                       | 82  |
| Figure A1.3 | Expression of MGAT1 after second siRNA transfection                      | 83  |
| Figure A1.4 | Cells infected by influenza after two siRNA transfections                | 84  |
| Figure A1.5 | Effect of cycloheximide on MGAT1   | 85  |
| Appendix 2  |  |     |
| Figure A2.1 | Three basic methods to link genotype and phenotype in directed evolution | 90  |
| Figure A2.2 | Crystal structure of Stoffel fragment polymerase                         | 92  |
| Figure A2.3 | Phage display method for DE of a polymerase                              | 94  |
| Figure A2.4 | Emulsion based method for DE of a polymerase                             | 95  |
| Figure A2.5 | Overview of novel polymerase DE method                                   | 96  |
| Figure A2.6 | Production of protein within an emulsion                                 | 98  |
| Figure A2.7 | Successful approach to SF polymerase selection                           | 99  |
| Figure A2.8 | Non-natural nucleotide incorporation by SF polymerase in emulsion        | 101 |
| Figure A2.9 | DNA obtained from first round of SF pol DE                               | 103 |

## LIST OF TABLES

### Chapter 1

|           |   |   |
|-----------|---|---|
| Table 1.1 | Classification of viral fusion proteins | 1 |
|-----------|---|---|

### Chapter 3

|           |  |    |
|-----------|--|----|
| Table 3.1 | Amino acid sequences of class I fusion peptides      | 44 |
| Table 3.2 | Sequence alignment of coronavirus S2' cleavage sites | 52 |
| Table 3.3 | Synthesized coronavirus fusion peptides              | 53 |

### Appendix 2

|            |   |     |
|------------|---|-----|
| Table A2.1 | Specialized primers for SF pol directed evolution               | 100 |
| Table A2.2 | Mutations of SF pol that allow incorporation of NTPs            | 102 |
| Table A2.3 | New mutations from SF pol DE with potential to incorporate NTPs | 104 |

# CHAPTER 1

## INTRODUCTION

### ***Viral Entry and Fusion***

The initial interactions of a virus and host cell that ultimately allow the completion of the virus replication cycle are commonly called virus entry [1]. These interactions typically begin with a binding event between the fusion protein of the virus and the surface of the host cell. Viral fusion proteins are critical to this initial stage of the virus life cycle because they are responsible for receptor binding as well as the merging of the viral lipid membrane and the host cell lipid membrane in the case of enveloped viruses.

Viral fusion proteins are divided into three classes based on their characteristics, as shown in Table 1.1.

Table 1.1 Classification of viral fusion proteins (adapted from Weissenhorn et. al, *FEBS Lett.*, 2007.)

| Virus Family            | Virus species                   | PDB code         |
|-------------------------|---------------------------------|------------------|
| <b><i>Class I</i></b>   |                                 |                  |
| Orthomyxoviridae        | Influenza virus A               | 1HA0, 3HMG, 1HTM |
| Paramyoviridae          | Human Parainfluenza virus F     | 1ZTM             |
|                         | Respiratory syncytial virus F   | 1G5G             |
| Filoviridae             | Ebola virus gp2                 | 1EBO, 2EBO       |
| Retroviridae            | HIV gp41                        | 1ENV, 1AIK       |
| Coronaviridae           | Mouse hepatitis virus S2        | 1WDG             |
| <b><i>Class II</i></b>  |                                 |                  |
| Flaviviridae            | Tick-borne encephalitis virus E | 1URZ, 1SVB       |
|                         | Dengue 2 and 3 virus E          | 2BEQ, 1WYY       |
| Togaviridae             | Semliki Forest virus E1         | 1OK8, 1UZG, 10AN |
| <b><i>Class III</i></b> |                                 |                  |
| Rhabdoviridae           | Vesicular stomatitis virus G    | 2CMZ, 2J6J       |
| Herpesviridae           | Herpes simplex virus gB         | 2GUM             |

Class I fusion proteins are typified by the influenza hemagglutinin protein.

Class I fusion proteins are composed of mainly alpha helical secondary structure, and usually form trimers [2-4]. They require proteolytic processing of the protein in a position close to the fusion peptide which leaves the protein in a metastable state prior to fusion. Fusion of class I fusion proteins can be triggered by receptor binding, low pH or a combination of the two. In contrast to class I fusion proteins, a typical class II fusion protein is the Semliki Forest virus E1/E2 protein[2-4]. Structurally, they are mainly composed of  $\beta$ -sheet like structures and exist as dimers in their native form. Although proteolytic processing is required for them to attain a fusion active state, an accessory protein is cleaved rather than the fusion protein itself. Additionally, low pH is the common mechanism of activation of class II fusion proteins. Recently a third class of fusion proteins has been established and notably contains, vesicular stomatitis virus (VSV) G protein and herpes simplex virus-1 (HSV-1) gB protein [4-6]. They are unique among fusion proteins due to their structure which is composed of roughly equal components of  $\alpha$ -helix and  $\beta$ -sheet, as well as other unique characteristics.

The first interaction between a virus and cell involves the binding of viral attachment proteins to specific cell surface receptors [7]. Viral receptors are defined as cell surfaced agents that directly bind with native virions and produce a biologic response [8]. Receptors can initiate conformational changes, activate signaling pathways, and/or promote endocytic internalization [1]. These receptors can be proteins, carbohydrates, or even lipids. It has also been found that the initial binding of the receptor to a virus can induce a conformational change which allows binding to a second receptor, usually called a co-receptor. The use of a co-receptor has been extensively studied for several viruses [9], most notably HIV [10]. A third and less specific type of interaction is also utilized by viruses. This interaction involves what is called an attachment factor which binds virus particles to the surfaced and concentrates them there [1]. Based on the importance of the receptor for the

initial interaction between the cell and the virus, it is easy to see that the distribution of the receptor on various tissues determines the cell tropism of the virus.

Subsequent to virus-receptor interaction, there are two possible methods of entry for the virus into the cell. In the case of paramyxoviruses and other pH independent viruses, entry continues with direct fusion of the viral lipid membrane with the plasma membrane of the host cell which deposits the nucleocapsid into the cytoplasm. Most RNA viruses replicate in the cytoplasm, but DNA viruses that fuse at the plasma membrane must replicate in the nucleus and their nucleocapsids are first transported before replication can begin. For influenza, VSV and other pH dependent viruses, receptor interactions are used to trigger endocytosis. Clathrin-mediated endocytosis is the most commonly used viral endocytic route [1], however clathrin-independent virus entry has also been identified as shown in Figure 1.1.

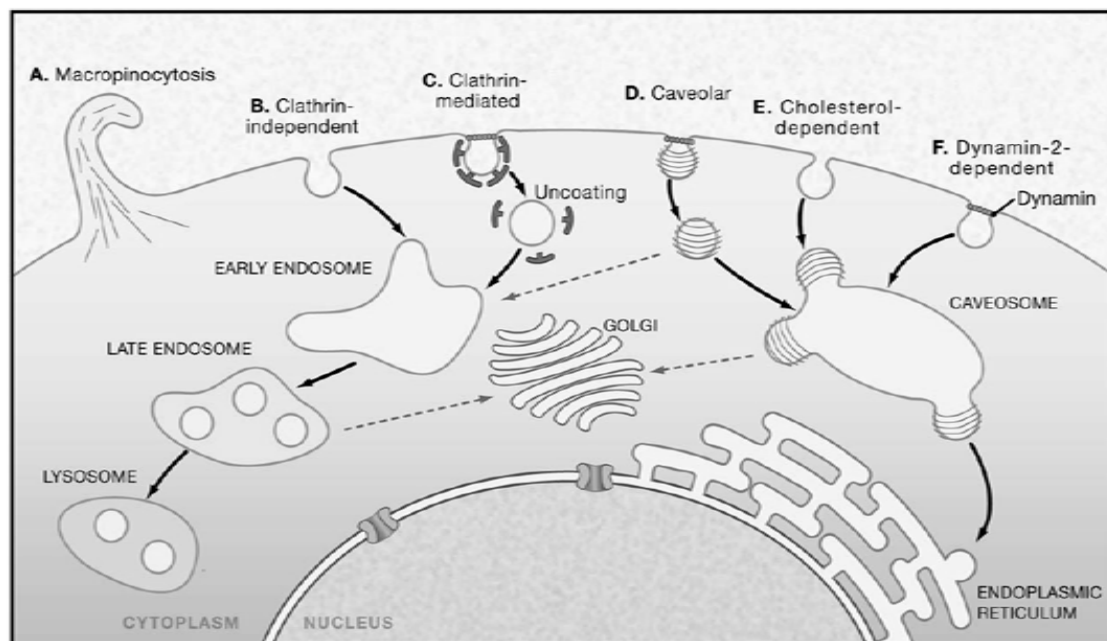


Figure 1.1 Entry mechanisms used by viruses (Marsh & Helenius, *Cell*, 2006.)

As mentioned previously, it is common for viruses and other pathogens to hijack the essential cell endocytic pathway to gain access to the cell cytoplasm. The

endocytic pathway normally serves to sort proteins and takes part in macromolecular/nutrient uptake/regulating signaling from receptors, and lysosome biogenesis [11]. The best understood mechanism of endocytosis is clathrin-mediated endocytosis [12], but other pathways are possible involving caveolae, polymorphous tubes and macropinosomes [13, 14]. In clathrin-mediated endocytosis, membrane proteins bind to cytosolic adaptors which are linked to the clathrin lattice. Additional accessory proteins are used to promote curvature of the membrane and clathrin polymerization. Finally, the vesicle is severed from the plasma membrane in a process thought to be mediated by dynamin [15]. There are several ways by which clathrin-mediated endocytosis can be triggered. One example involves short linear amino acid sequences present on the cytosolic tail of proteins such as YXXΦ and [DE]xxxLL motifs which can be recognized AP2, a clathrin adaptor protein [16]. Another mechanism involves protein modification by ubiquitin which is recognized by the EPS15 and epsin clathrin adaptor proteins [17]. After uptake from the plasma membrane, a protein will arrive at an early endosome, followed by a late endosome, and ultimately be delivered to the trans-Golgi network (TGN), a lysosome or recycled back to the plasma membrane. Efficient retrograde transport from endosomes to the TGN is limited to a specific set of proteins, of which the most is known about the acid-hydrolase receptors, transmembrane enzymes and SNAREs [18]. Although the retrograde-transport mechanism for most proteins is unknown, some information has been established regarding the transport of acid-hydrolase receptors. It is believed that the process initiates with the budding of transport carriers from early or late endosomes involving a protein complex named a retromer [18], which travels to the TGN where it fuses. It is possible that transport to the TGN of other proteins also involves components of this system. Proteins are also transported from the early endosome to the lysosome for destruction. The process of sorting proteins to the

lysosome begins in the late endosome. Late endosomes contain intraluminal vesicles which are formed by the invagination of the limiting membrane of the endosome. The sorting of cargo into these intraluminal vesicles, which is signaled by the ubiquitination of these proteins, dooms them to be degraded in the lysosome [12]. Numerous hypotheses have been proposed regarding how proteins are transferred from late endosomes to the lysosome [12]. Although the kiss-and-run [19] and hybrid compartment models [20, 21] have the most experimental evidence behind them, more work must be done to conclusively determine this mechanism. Finally, proteins can be taken up from the plasma membrane by the endocytic pathway and returned. This method of transport is necessary to maintain a balance of receptors at the cell surface and aid in absorption of macromolecules [11]. For many proteins “recycled” in this way, the mechanism of sorting is not fully understood. One method used by the LDL receptor related protein LRP is the presence of a cytoplasmic tail motif NPxY which interacts with the sorting nexin protein SNX17 [11]. Another widely known protein the transferrin receptor (TfnR) is also recycled from the plasma membrane. It is thought to use “fast” recycling pathway in the early endosome [22]. From the complexity of the endosomal system and the vast variety of ways in which it is utilized, it is easy to see how it would be advantageous for pathogens to “borrow” the system.

At some point in the endocytic pathway, fusion between the viral lipid membrane and the endosomal membrane is triggered and the nucleocapsid is deposited into the cytoplasm. Fusion and fission events also regularly occur in the cell under routine circumstances. SNARE (SNAP receptors) proteins are the central mediators of these fusion events which are exemplified in the prior discussion of the endocytic pathway. SNAREs bind SNAP (soluble NSF attachment proteins) and NSF (N-ethylmaleimide sensitive factor) [23].



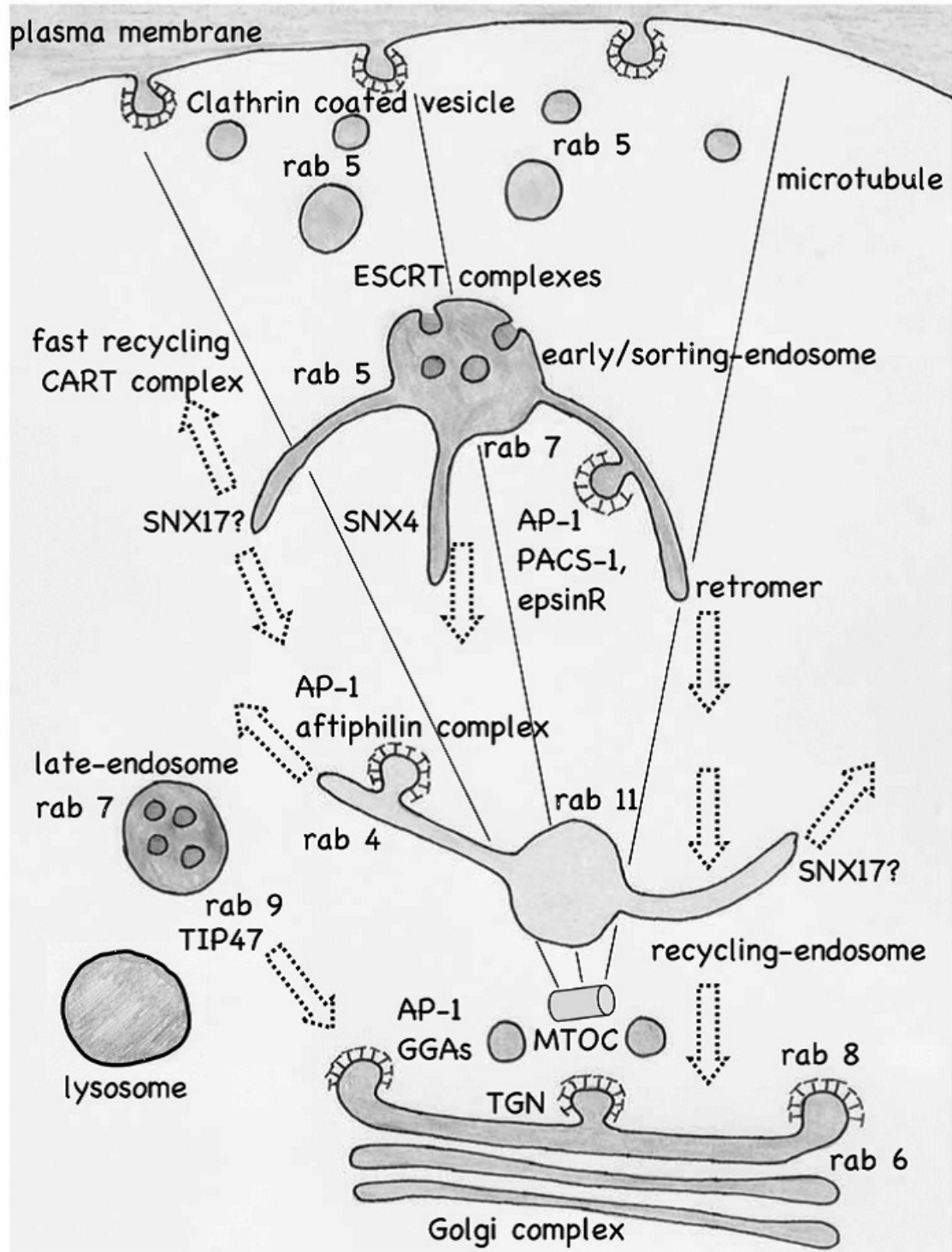


Figure 1.2 A simplified diagram of the endocytic pathway (Seaman. *Cell. Mol. Life Sci.*, 2008.)

SNAREs are typically type II membrane proteins of which the majority of the protein is exposed to the cytoplasm. SNAREs are divided into two groups, the v-SNAREs which are found on the vesicles being transported and t-SNAREs which are found on the target membrane [24]. Although there are many SNAREs and therefore a large number of combinations of SNAREs that could form complexes, only a small number can actually promote fusion. In this way, SNAREs can provide specificity to the fusion reaction. The initial step in vesicle-target fusion is an interaction involving a tethering protein and a Rab protein [25]. This interaction brings the SNARE proteins present on the two surfaces close enough together to form a SNARE-pin which consists of an assembly of parallel four-helix bundle of the SNARE protein and an SM (Sec1/Munc18-like) protein. The “zipping up” of the four-helix bundle provides enough energy to pull the two membranes close enough to drive fusion. One SNARE pin is sufficient to achieve hemifusion but at least three are required for full fusion [26]. After fusion has occurred, the complex ends up in a configuration called the cis-SNARE complex, which can be reactivated in an ATP dependent manner with NSF. Although the details of the mechanism are different, SNARE complexes and viral fusion proteins act in a similar manner to destabilize membranes and bring them close enough together to promote a fusion event [27].

The key component of the fusion protein which enables fusion is the fusion peptide [28]. The fusion peptide is the portion of the fusion protein that is inserted into the host cell membrane. Not only does the insertion of the fusion peptide into the host cell membrane serve to bring the two membranes close enough to promote mixing, but they also alter the organization and structure of the lipid bilayers they are inserted into [29].

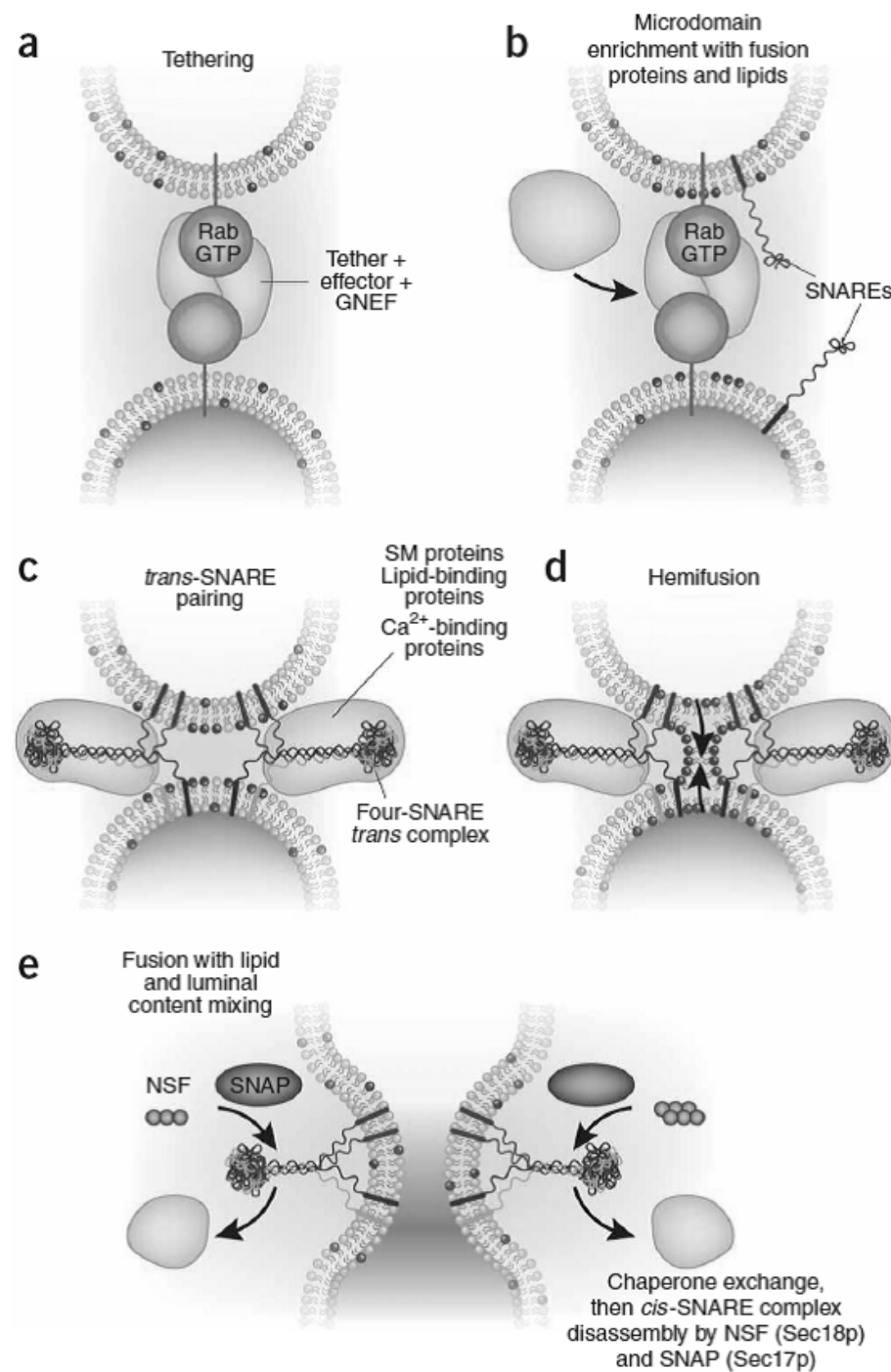


Figure 1.3 Membrane fusion in the endocytic pathway, steps a-e (Wickner & Schekman, *Nat. Struct. Mol. Biol.*, 2008.)

The work contained in this thesis examines several aspects of entry,

both receptor binding and fusion, for several viruses. In chapter 2, a line of research examining the effect of membrane lipid composition on VSV fusion is examined. Characterization of a novel coronavirus fusion peptide is the topic of chapter 3. Finally, appendix 1 examines possible receptor interactions of influenza. In the remainder of this introduction, I will attempt to lay a foundation for the understanding of these studies by providing a brief overview of what is currently known about the process of virus entry for each of the studied viruses.

### ***Vesicular Stomatitis Virus (VSV)***

Vesicular stomatitis virus (VSV), Figure 1.4, is a member of the *Rhabdoviridae* [30]. As such, it is an enveloped virus and has a non-segmented, negative stranded RNA genome which encodes 5 proteins including a glycoprotein (G), matrix protein (M), phosphoprotein (P), large protein (L) and a nucleoprotein (N). Of key importance to VSV entry and fusion is the G protein which is responsible for receptor binding and fusion of the viral and host membrane which results in the delivery of the nucleocapsid to the cell cytoplasm.

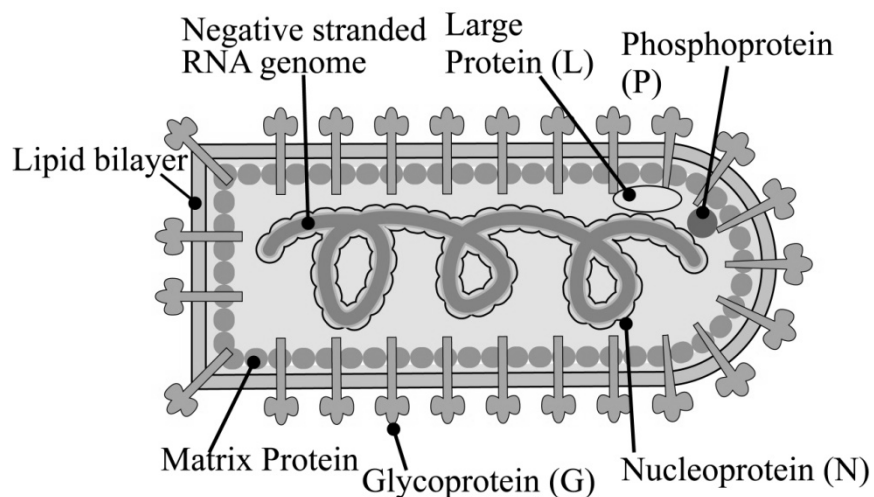


Figure 1.4 Schematic representation of Vesicular stomatitis virus (adapted from Dr. Ruth Collins with permission, personal communication).

The X-ray crystal structures of the VSV G protein in its pre and post fusion forms were recently elucidated [31, 32]. This structure afforded many insights into the mechanism of the G protein and also established it as a member of a new class of fusion proteins as discussed previously[33]. One important revelation was the identification of bipartite fusion loops in contrast to a fusion peptide. Mutagenesis experiments performed on the residues in these fusion loops show they are intolerant to manipulation [34]. Although the crystal structure has shed light into the mechanisms of G protein rearrangement upon exposure to low pH and the active fusion domain, the VSV receptor binding site and the identity of the receptor itself are still unknown.

An early attempt at the identification of the VSV receptor showed that cell membrane extracts were no longer inhibitory to VSV binding after being treated with phospholipase C but not after treatment with proteases, neuraminidase or heat [35]. This initial observation led to the hypothesis that the receptor for VSV was a lipid. Additionally, the same investigators were able to show that when VSV was exposed to phosphatidylserine or liposomes containing phosphatidylserine VSV binding to cell surfaces was inhibited. Since those observations, it has been observed that VSV G protein interacts with several other lipids besides phosphatidylserine. Some initial studies used purified lipid and carbohydrate components to inhibit virus promoted cell-cell fusion [36]. In addition to phosphatidylserine, phosphatidylinositol, sphingomyelin, cholesterol, and the ganglioside GM3 were able to inhibit VSV promoted cell-cell fusion. Subsequent biochemical experiments utilizing isothermal titration calorimetry and atomic force microscopy confirmed the interaction between VSV G and phosphatidylserine and also suggested an interaction with cardiolipin and phosphatidylglycerol [37, 38]. In addition to the VSV G protein's tendency to non-specifically interact with several different lipids, other information has surfaced

refuting the hypothesis that phosphatidylserine is the receptor for VSV. It has been shown that there is no correlation between the amount of phosphatidylserine in the plasma membrane of a cell and the degree of VSV binding [39]. Also, phosphatidylserine is mostly present on the internal leaflet of the cell membrane and therefore not readily available to act as a cell surface receptor [40]. Due to the current confusion over this topic, further work to identify the receptor for VSV is clearly needed.

After the receptor binding event, VSV enters the endocytic pathway in a clathrin dependent fashion [41-44]. The exact process of fusion and infection, however, is less clear. In 2005, Le Blanc et al proposed a two step mechanisms for VSV fusion and infection [45]. Using time-lapse confocal microscopy and a pH sensitive dye they determined that fusion must occur at a later stage than the early endosome. Using inhibitors like nocodazole, they determined that fusion must occur before VSV reached the late endosome. This evidence pointed to fusion of the VSV virion into the internal vesicles of the endosomal carrier vesicles or multivesicular bodies (ECV/MVB). On top of that, they were able to capture EM images showing VSV nucleocapsids inside of the internal vesicles of multivesicular bodies. Regarding VSV infection, they observed that it was disrupted by agents that disrupted late endosomes such as nocodazole (in contrast to fusion) and that inhibiting the back fusion of the internal vesicles of the MVBs with an anti-LBPA antibody also inhibited infection. In total, they proposed a fusion step of VSV with the internal vesicles of the MVBs followed by infection by VSV as a result of back fusion of those vesicles with the limiting membrane of the late endosome. In early 2009, a paper was published that proposed an alternative pathway for VSV fusion and infection [42]. Using time-lapse confocal microscopy and EM they determined that fusion and infection occurred in one step in the early endosome by fusion of the endosomal membrane with the viral

lipid membrane which resulted in the introduction of the nucleocapsid into the cell cytoplasm. A comparison of the two models can be seen in Figure 1.5. One key difference between these two models is the conditions under which fusion occurs. In the two step model, VSV is fusing to a membrane that contains a specific lipid found only in the membrane of the internal vesicles, lysobisphosphatidic acid (LBPA). In chapter 2, I provide evidence that this lipid interacts with the VSV G protein in a manner that accelerates fusion, which could contribute to the phenomenon of the two step fusion-infection mechanism.

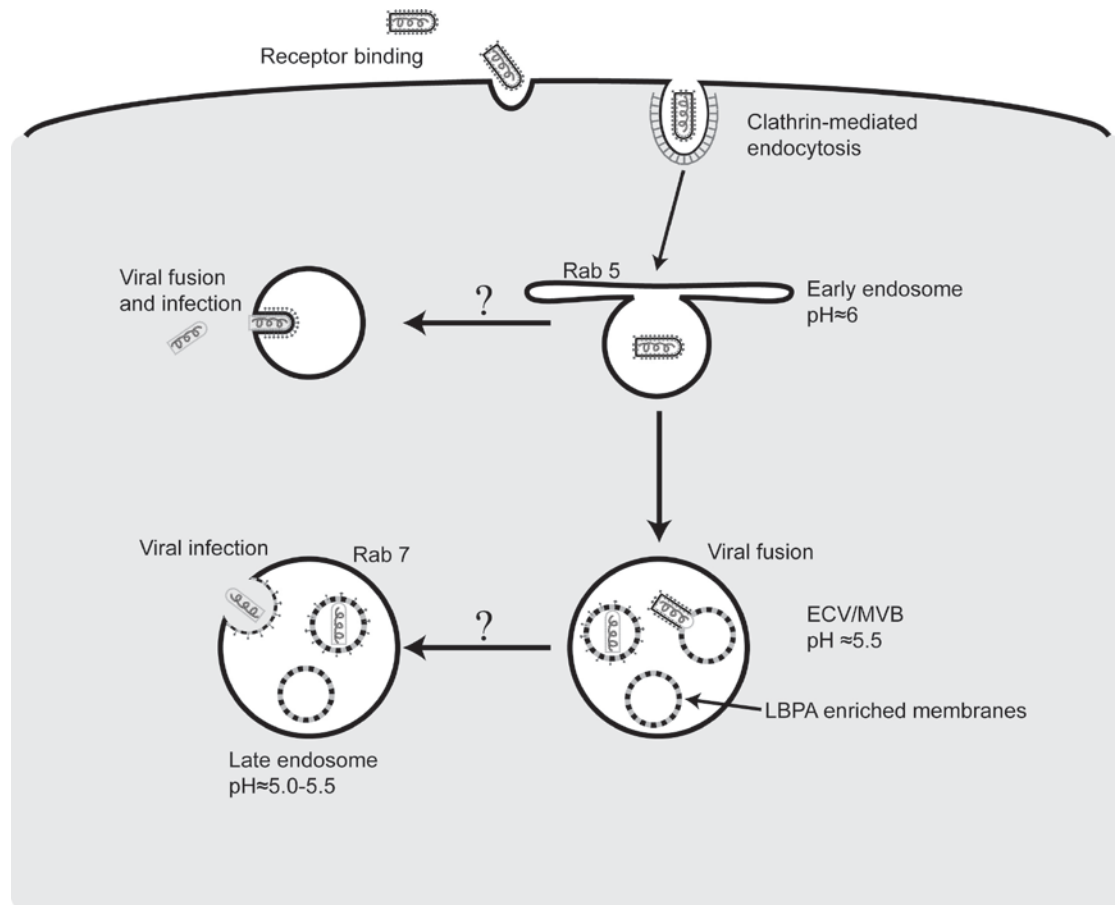


Figure 1.5 The two proposed VSV endocytic entry pathways.

### ***Coronaviruses***

The *Coronaviridae* are large, enveloped positive stranded RNA viruses that

can infect humans as well as other domestic birds and mammals [46], Figure 1.6. The family is customarily split into three groups. Of particular interest here are the group I feline coronaviruses, feline enteric coronavirus (FeCoV) and feline infectious peritonitis virus (FIPV). Among group II viruses, severe acute respiratory syndrome corona virus (SARS-CoV) a human pathogen, which is actually classified into a subgroup of group II, will be a subject of further investigation. Finally, among group III viruses, infectious bronchitis virus (IBV) a significant pathogen of chickens will be discussed in comparison with SARS and the feline coronaviruses. Of particular importance to entry and fusion of coronaviruses is the spike (S) glycoprotein which is a large, heavily glycosylated protein of approximately 150-180 kDa which is embedded into the viral lipid envelope [46].

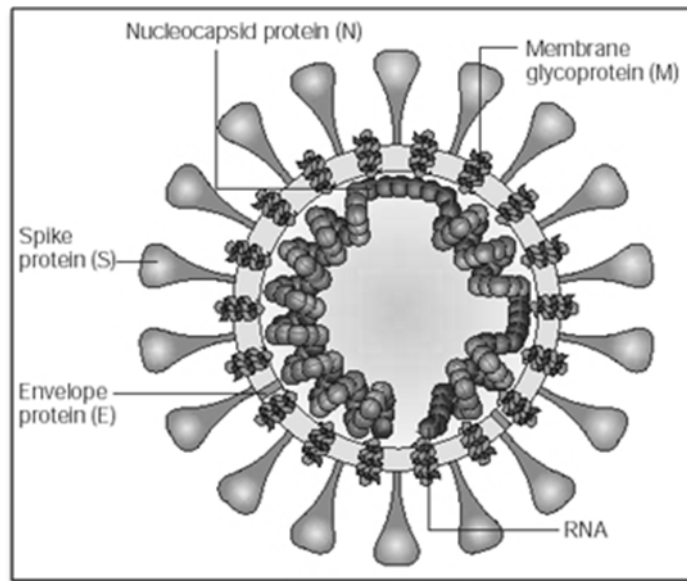


Figure 1.6 Schematic representation of coronavirus structure (Stadler, *Nat. Micro.*, 2003.)

The coronavirus spike protein is a class I fusion protein [47], however there seems to be some variation of certain characteristics such as the necessity for



proteolytic cleavage amongst the coronavirus spike proteins. Although no X-ray crystal structure has been obtained of the SARS-CoV S protein, there has been a concerted effort to determine aspects of its structure and mechanism of action [48]. Some structural information has been gleaned from a crystal structure of the receptor binding domain complexed to the receptor angiotensin converting enzyme-2 (ACE-2) [49]. Further, studies of the heptad repeats have shown that the protein likely exists as a monomer in the pre-fusion state and that, upon activation, the heptad repeat domains interact in a six-helix bundle and the protein associates into a homotrimer [47, 50]. A number of studies of the mechanism of the SARS-CoV S protein have attempted to understand the proteolytic processing of the protein. Studies have shown that the spike protein can be cleaved by factor Xa [51], but more importantly, cathepsin L [52, 53] during entry by the endosomal pathway, and trypsin promoting fusion at the cell surface [54-56]. A more recent study shows that a sequential cleavage event is necessary to fully activate the SARS-CoV S protein [57].



Figure 1.7 Schematic representation of coronavirus spike protein. The protein is divided into two sections, S1 and S2. The two heptad repeats are labeled as HR1 and HR2 and the transmembrane domain is labeled TM. The cleavage site internal to S2 is labeled S2'.

As mentioned previously, proteolytic cleavage of class I fusion proteins commonly exposes the fusion peptide. There is currently a controversy about the identity of the SARS-CoV fusion peptide in the literature. Several sequences in the protein have

been characterized as having the features of a fusion peptide. Initial reports used the Wimley and White interfacial hydrophobicity scale to identify two regions including amino acids M770-L778 (WW-I) and A864-Y886 as possible fusion peptides. These peptides were shown to induce fusion and leakage of lipid vesicles and possess secondary structure alone and in the presence of lipid vesicles [58]. The same group has also proposed amino acids 858-886 could be the SARS-CoV fusion peptide in analogy to the fusion peptide of the HIV gp41 glycoprotein [59-61]. In chapter 3, I will discuss my work characterizing our proposed fusion peptide for SARS-CoV.

Also presented in chapter 3 is the characterization of the analogous fusion peptide for the feline coronaviruses. The feline coronaviruses are divided into two serotypes, type I which account for approximately 80% of all infections [62, 63], and type II which are more widely studied in the laboratory due to their ability to be propagated in cell culture. Each serotype can be further divided into two biotypes, feline enteric coronavirus (FeCoV) and feline infectious peritonitis virus (FIPV) which can only be distinguished biologically [64]. FeCoV is a mild enteric infection which in less than 5% of cats, can mutate into the fatal systemic biotype FIPV [65]. Mutations have been identified in the 3c gene, which encodes for a protein of unknown function, that distinguish FIPV from FECV [66]. However, the “internal mutation” of FECV into FIPV remains controversial and alternative theories have been proposed [67]. It has been well established that type I feline coronavirus spike glycoproteins use aminopeptidase N (APN) as their receptor [68]. A recent report shows that DC-SIGN can enhance the infection of type II feline coronaviruses in permissive cells and rescue it in non-permissive cells [69]. However, the receptor for serotype I viruses is currently unknown [70, 71]. A recent report by Regan et. al shows that FECV is sensitive to pH as well as cathepsin L and cathepsin B inhibitors and FIPV is insensitive to pH and cathepsin L inhibitors but sensitive to cathepsin B

inhibitors. This data could point to different routes of entry for these two viruses, FECV utilizing the endosomal pathway and FIPV being able to fuse at the cell surface. It also infers that the spike protein of FECV must be cleaved by cathepsins L and B while the spike protein of FIPV must only be cleaved by cathepsin B, although further work must be done to confirm these hypotheses. In chapter 3, I will discuss the evidence that suggests these two coronaviruses have slightly different fusion peptides and the effect that has on their ability to promote membrane fusion.

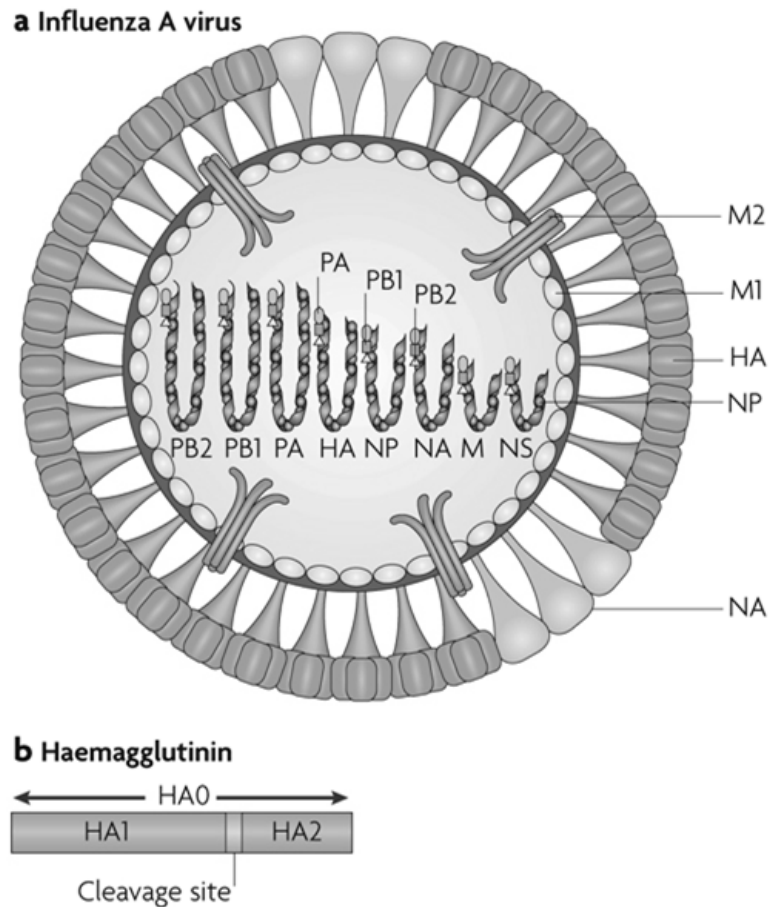


Figure 1.8 Schematic representation of Influenza A virus. A. Influenza virion, B. Hemagglutinin (HA) protein (Subbarao & Joseph, *Nat. Rev. Immunol.*, 2007).

## ***Influenza***

Influenza is an enveloped virus with a negative single-stranded RNA genome, as shown in Figure 1.8 [72]. There are 8 segments to the genome which encode 10 or 11 proteins [73]. Much is known about influenza viruses because they are one of the main model systems for host cell-virus interactions. Of particular interest to studies of entry is the influenza glycoprotein hemagglutinin (HA) which is responsible for receptor binding and entry. Influenza hemagglutinin is a class I fusion protein. Further discussion of the receptor binding activities of HA can be found in Appendix 1. Prior to cleavage, HA is referred to as HA0, the protein is cleaved into two sulfide bridged pieces, HA1 which includes the receptor binding domain, and HA2 which is responsible for fusion and contains the fusion peptide [4]. To allow for fusion, the low pH environment of the endosome is necessary to promote structural rearrangement of the HA protein and exposure of the fusion peptide. Influenza enters the cell using multiple endocytic pathways. Single particle experiments show that influenza uses clathrin-mediated endocytosis approximately 60% of the time but also uses clathrin and caveolin independent pathways [74]. Studies of the intracellular trafficking of influenza virus in polar cells show a significant role of the actin cytoskeleton in internalization [75]. In appendix 1, I present my work on defining the role of N-linked glycosylation on the entry of influenza virus into human cells.

## REFERENCES

1. Marsh, M. and A. Helenius, *Virus entry: Open sesame*. Cell, 2006. **124**(4): p. 729-740.
2. Kielian, M. and F.A. Rey, *Virus membrane-fusion proteins: more than one way to make a hairpin*. Nature Reviews Microbiology, 2006. **4**(1): p. 67-76.
3. Schibli, D.J. and W. Weissenhorn, *Class I and class II viral fusion protein structures reveal similar principles in membrane fusion (Review)*. Molecular Membrane Biology, 2004. **21**(6): p. 361-371.
4. White, J.M., et al., *Structures and mechanisms of viral membrane fusion proteins: Multiple variations on a common theme (vol 43, pg 189, 2008)*. Critical Reviews in Biochemistry and Molecular Biology, 2008. **43**(4): p. 287-288.
5. Harrison, S.C., *Viral membrane fusion*. Nature Structural & Molecular Biology, 2008. **15**(7): p. 690-698.
6. Weissenhorn, W., A. Hinz, and Y. Gaudin, *Virus membrane fusion*. Febs Letters, 2007. **581**(11): p. 2150-2155.
7. Young, J.A.T., *Virus entry and uncoating*, in *Fields Virology*, D.M. Knipe and P.M. Howley, Editors. 2001, Lippincott Williams & Wilkins: Philadelphia, PA. p. 87-104.
8. Lentz, T.L., *The recognition event between virus and host-cell receptor - a target for antiviral agents*. Journal of General Virology, 1990. **71**: p. 751-766.
9. Haywood, A.M., *Virus receptors - binding, adhesion strengthening, and changes in viral structure*. Journal of Virology, 1994. **68**(1): p. 1-5.
10. Dragic, T., *An overview of the determinants of CCR5 and CXCR4 co-receptor function*. Journal of General Virology, 2001. **82**: p. 1807-1814.
11. Seaman, M.N.J., *Endosome protein sorting: motifs and machinery*. Cellular

- and Molecular Life Sciences, 2008. **65**(18): p. 2842-2858.
12. Pryor, P.R. and J.P. Luzio, *Delivery of endocytosed membrane proteins to the lysosome*. Biochimica Et Biophysica Acta-Molecular Cell Research, 2009. **1793**(4): p. 615-624.
  13. Donaldson, J.G., N. Porat-Shliom, and L.A. Cohen, *Clathrin-independent endocytosis: A unique platform for cell signaling and PM remodeling*. Cellular Signalling, 2009. **21**(1): p. 1-6.
  14. Hansen, C.G. and B.J. Nichols, *Molecular mechanisms of clathrin-independent endocytosis*. Journal of Cell Science, 2009. **122**(11): p. 1713-1721.
  15. Rappoport, J.Z., *Focusing on clathrin-mediated endocytosis*. Biochemical Journal, 2008. **412**: p. 415-423.
  16. Bonifacino, J.S. and L.M. Traub, *Signals for sorting of transmembrane proteins to endosomes and lysosomes*. Annual Review of Biochemistry, 2003. **72**: p. 395-447.
  17. Piper, R.C. and J.P. Luzio, *Ubiquitin-dependent sorting of integral membrane proteins for degradation in lysosomes*. Current Opinion in Cell Biology, 2007. **19**(4): p. 459-465.
  18. Bonifacino, J.S. and R. Rojas, *Retrograde transport from endosomes to the trans-Golgi network*. Nature Reviews Molecular Cell Biology, 2006. **7**(8): p. 568-579.
  19. Bright, N.A., M.J. Gratian, and J.P. Luzio, *Endocytic delivery to lysosomes mediated by concurrent fusion and kissing events in living cells*. Current Biology, 2005. **15**(4): p. 360-365.
  20. Bright, N.A., et al., *Dense core lysosomes can fuse with late endosomes and are re-formed from the resultant hybrid organelles*. Journal of Cell Science, 1997. **110**: p. 2027-2040.

21. Mullock, B.M., et al., *Fusion of lysosomes with late endosomes produces a hybrid organelle of intermediate density and is NSF dependent*. Journal of Cell Biology, 1998. **140**(3): p. 591-601.
22. Sheff, D., et al., *Transferrin receptor recycling in the absence of perinuclear recycling endosomes*. Journal of Cell Biology, 2002. **156**(5): p. 797-804.
23. Malsam, J., S. Kreye, and T.H. Sollner, *Membrane fusion: SNAREs and regulation*. Cellular and Molecular Life Sciences, 2008. **65**(18): p. 2814-2832.
24. Ungermann, C. and D. Langosch, *Functions of SNAREs in intracellular membrane fusion and lipid bilayer mixing*. Journal of Cell Science, 2005. **118**(17): p. 3819-3828.
25. Wickner, W. and R. Schekman, *Membrane fusion*. Nature Structural & Molecular Biology, 2008. **15**(7): p. 658-664.
26. Sudhof, T.C. and J.E. Rothman, *Membrane Fusion: Grappling with SNARE and SM Proteins*. Science, 2009. **323**(5913): p. 474-477.
27. Martens, S. and H.T. McMahon, *Mechanisms of membrane fusion: disparate players and common principles*. Nature Reviews Molecular Cell Biology, 2008. **9**(7): p. 543-556.
28. Durell, S.R., et al., *What studies of fusion peptides tell us about viral envelope glycoprotein-mediated membrane fusion*. Molecular Membrane Biology, 1997. **14**(3): p. 97-112.
29. Tamm, L.K., et al., *Structure and function of membrane fusion peptides*. Biopolymers, 2002. **66**(4): p. 249-260.
30. Rose, J.K., and Whitt, M. A., *Rhabdoviridae: The viruses and their replication*, in *Fields Virology*, D.M. Knipe and P.M. Howley, Editors. 2001, Lippincott Williams & Wilkins: Philadelphia, PA. p. 1221-1244.
31. Roche, S., et al., *Crystal structure of the low-pH form of the vesicular*

- stomatitis virus glycoprotein G*. Science, 2006. **313**(5784): p. 187-191.
32. Roche, S., et al., *Structure of the prefusion form of the vesicular stomatitis virus glycoprotein G*. Science, 2007. **315**(5813): p. 843-848.
  33. Roche, S., et al., *Structures of vesicular stomatitis virus glycoprotein: membrane fusion revisited*. Cellular and Molecular Life Sciences, 2008. **65**(11): p. 1716-1728.
  34. Sun, X., S. Belouzard, and G.R. Whittaker, *Molecular architecture of the bipartite fusion loops of vesicular stomatitis virus glycoprotein G, a class III viral fusion protein*. Journal of Biological Chemistry, 2008. **283**(10): p. 6418-6427.
  35. Schlegel, R., et al., *Inhibition of VSV binding and infectivity by phosphatidylserine - is phosphatidylserine a VSV-binding site*. Cell, 1983. **32**(2): p. 639-646.
  36. Conti, C., P. Mastromarino, and N. Orsi, *Role of membrane phospholipids and glycolipids in cell-to-cell fusion by VSV*. Comparative Immunology Microbiology and Infectious Diseases, 1991. **14**(4): p. 303-&.
  37. Carneiro, F.A., et al., *Membrane recognition by vesicular stomatitis virus involves enthalpy-driven protein-lipid interactions*. Journal of Virology, 2002. **76**(8): p. 3756-3764.
  38. Carneiro, F.A., et al., *Probing the interaction between vesicular stomatitis virus and phosphatidylserine*. European Biophysics Journal with Biophysics Letters, 2006. **35**(2): p. 145-154.
  39. Coil, D.A. and A.D. Miller, *Phosphatidylserine is not the cell surface receptor for vesicular stomatitis virus*. Journal of Virology, 2004. **78**(20): p. 10920-10926.
  40. Boon, J.M. and B.D. Smith, *Chemical control of phospholipid distribution*



- across bilayer membranes*. Medicinal Research Reviews, 2002. **22**(3): p. 251-281.
41. Cureton, D.K., et al., *Vesicular Stomatitis Virus Enters Cells through Vesicles Incompletely Coated with Clathrin That Depend upon Actin for Internalization*. Plos Pathogens, 2009. **5**(4).
  42. Johannsdottir, H.K., et al., *Host Cell Factors and Functions Involved in Vesicular Stomatitis Virus Entry*. Journal of Virology, 2009. **83**(1): p. 440-453.
  43. Sun, X.J., et al., *Role of clathrin-mediated endocytosis during vesicular stomatitis virus entry into host cells*. Virology, 2005. **338**(1): p. 53-60.
  44. Superti, F., et al., *Entry pathway of vesicular stomatitis-virus into different host-cells*. Journal of General Virology, 1987. **68**: p. 387-399.
  45. Le Blanc, I., et al., *Endosome-to-cytosol transport of viral nucleocapsids*. Nature Cell Biology, 2005. **7**(7): p. 653-U25.
  46. Lai, M.M.C. and K.V. Holmes, *Coronaviridae: The viruses and their replication*, in *Fields Virology*, D.M. Knipe and P.M. Howley, Editors. 2001, Lippincott Williams and Wilkins: Philadelphia, PA. p. 1163-1186.
  47. Bosch, B.J., et al., *The coronavirus spike protein is a class I virus fusion protein: Structural and functional characterization of the fusion core complex*. Journal of Virology, 2003. **77**(16): p. 8801-8811.
  48. Du, L.Y., et al., *The spike protein of SARS-CoV - a target for vaccine and therapeutic development*. Nature Reviews Microbiology, 2009. **7**(3): p. 226-236.
  49. Li, F., et al., *Structure of SARS coronavirus spike receptor-binding domain complexed with receptor*. Science, 2005. **309**(5742): p. 1864-1868.
  50. McReynolds, S., et al., *Characterization of the prefusion and transition states of severe acute respiratory syndrome coronavirus S2-HR2*. Biochemistry,

2008. **47**(26): p. 6802-6808.
51. Du, L.Y., et al., *Cleavage of spike protein of SARS coronavirus by protease factor Xa is associated with viral infectivity*. Biochemical and Biophysical Research Communications, 2007. **359**(1): p. 174-179.
  52. Bosch, B.J., W. Bartelink, and P.J.M. Rottier, *Cathepsin L functionally cleaves the severe acute respiratory syndrome coronavirus class I fusion protein upstream of rather than adjacent to the fusion peptide*. Journal of Virology, 2008. **82**(17): p. 8887-8890.
  53. Simmons, G., et al., *Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry*. Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(33): p. 11876-11881.
  54. de Haan, C.A.M., et al., *Cleavage of group 1 coronavirus spike proteins: How furin cleavage is traded off against heparan sulfate binding upon cell culture adaptation*. Journal of Virology, 2008. **82**(12): p. 6078-6083.
  55. Li, F., et al., *Conformational states of the severe acute respiratory syndrome coronavirus spike protein ectodomain*. Journal of Virology, 2006. **80**(14): p. 6794-6800.
  56. Matsuyama, S., et al., *Protease-mediated enhancement of severe acute respiratory syndrome coronavirus infection*. Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(35): p. 12543-12547.
  57. Belouzard, S., V.C. Chu, and G.R. Whittaker, *Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites*. Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(14): p. 5871-5876.
  58. Sainz, B., et al., *Identification and characterization of the putative fusion*

- peptide of the severe acute respiratory syndrome-associated coronavirus spike protein*. Journal of Virology, 2005. **79**(11): p. 7195-7206.
59. Guillen, J., et al., *Structural and dynamic characterization of the interaction of the putative fusion peptide of the S2SARS-CoV virus protein with lipid membranes*. Journal of Physical Chemistry B, 2008. **112**(23): p. 6997-7007.
  60. Guillen, J., P.K.J. Kinnunen, and J. Villalain, *Membrane insertion of the three main membranotropic sequences from SARS-CoV S2 glycoprotein*. Biochimica Et Biophysica Acta-Biomembranes, 2008. **1778**(12): p. 2765-2774.
  61. Guillen, J., et al., *A second SARS-CoV S2 glycoprotein internal membrane-active peptide. Biophysical characterization and membrane interaction*. Biochemistry, 2008. **47**(31): p. 8214-8224.
  62. Addie, D.D., et al., *Persistence and transmission of natural type I feline coronavirus infection*. Journal of General Virology, 2003. **84**: p. 2735-2744.
  63. Hohdatsu, T., et al., *The prevalence of type-I and type-II feline coronavirus infections in cats*. Journal of Veterinary Medical Science, 1992. **54**(3): p. 557-562.
  64. Vennema, H. *Genetic drift and genetic shift during feline coronavirus evolution*. 1999: Elsevier Science Bv.
  65. Pedersen, N.C. *An overview of feline enteric coronavirus and infectious peritonitis virus-infections*. 1995: Veterinary Practice Publ Co.
  66. Pedersen, N.C., *A review of feline infectious peritonitis virus infection: 1963-2008*. Journal of Feline Medicine and Surgery, 2009. **11**(4): p. 225-258.
  67. Dye, C. and S.G. Siddell, *Genomic RNA sequence of feline coronavirus strain FCoVCIJe*. Journal of Feline Medicine and Surgery, 2007. **9**(3): p. 202-213.
  68. Tresnan, D.B., R. Levis, and K.V. Holmes, *Feline aminopeptidase N serves as a receptor for feline, canine, porcine, and human coronaviruses in serogroup*

- I.* Journal of Virology, 1996. **70**(12): p. 8669-8674.
69. Regan, A.D. and G.R. Whittaker, *Utilization of DC-SIGN for Entry of Feline Coronaviruses into Host Cells*. Journal of Virology, 2008. **82**(23): p. 11992-11996.
  70. Dye, C., N. Temperton, and S.G. Siddell, *Type I feline coronavirus spike glycoprotein fails to recognize aminopeptidase N as a functional receptor on feline cell lines*. Journal of General Virology, 2007. **88**: p. 1753-1760.
  71. Hohdatsu, T., et al., *Differences in virus receptor for type I and type II feline infectious peritonitis virus*. Archives of Virology, 1998. **143**(5): p. 839-850.
  72. Subbarao, K. and T. Joseph, *Scientific barriers to developing vaccines against avian influenza viruses*. Nature Reviews Immunology, 2007. **7**(4): p. 267-278.
  73. Lamb, R.A. and R.M. Krug, *Orthomyxoviridae: The viruses and their replication*, in *Fields Virology*, D.M. Knipe and P.M. Howley, Editors. 2001, Lippincott Williams & Wilkins: Philadelphia. p. 1487-1532.
  74. Rust, M.J., et al., *Assembly of endocytic machinery around individual influenza viruses during viral entry*. Nature Structural & Molecular Biology, 2004. **11**(6): p. 567-573.
  75. Sun, X.J. and G.R. Whittaker, *Role of the actin cytoskeleton during influenza virus internalization into polarized epithelial cells*. Cellular Microbiology, 2007. **9**(7): p. 1672-1682.

## CHAPTER 2

### PROMOTION OF VIRAL MEMBRANE FUSION BY THE ENDOSOME-SPECIFIC PHOSPHOLIPID LYSOBISPHOSPHATIDIC ACID (LBPA)

#### ***Abstract***

Vesicular stomatitis virus (VSV) is a prototypic virus that is commonly used in studies of endocytosis and membrane trafficking. A proposed mechanism for VSV entry involves virus internalization into endosomes, followed by a two-step fusion reaction; initial fusion with internal vesicles of multivesicular endosomes followed by subsequent back-fusion of these vesicles into the cytoplasm. One feature of the internal vesicles of late endosomes is that they uniquely contain the lipid lysobisphosphatidic acid (LBPA). Using a FRET-based *in vitro* lipid mixing assay, we show that the presence of LBPA significantly increases the rate of VSV G mediated membrane fusion. The increased rate of lipid mixing was selective for VSV and was not evident when other viruses, such as influenza, were examined. Our data provide a biological rationale for a two-step fusion reaction during VSV entry, and suggests that LBPA may preferentially affect the ability of VSV G (a class III viral fusion protein) to mediate lipid mixing during membrane fusion.

#### ***Introduction***

The endocytic pathway is responsible for the maintenance of the cell surface composition and is usurped by many viruses as a portal of entry into the cell [1]. Such viruses often take advantage of the low endosomal pH to trigger their uncoating and/or fusion and thus gain access to the cytosol for replication. While low pH is well recognized as a way to regulate the conformational changes occurring during virus entry, the role of other endosomal factors such as lipid composition remains poorly characterized. Vesicular stomatitis virus (VSV) enters cells via clathrin-mediated

endocytosis [2, 3], with entry into the cytosol mediated via the viral glycoprotein G [4]. VSV G mediates membrane fusion in a low pH-dependent manner, with optimal fusion activity at pH 5.7–5.8 or below [3]. VSV G shows unusual properties for a viral fusion protein, in that the conformational changes triggering membrane fusion are reversible [5]. Another unusual feature of VSV is that viral entry has been proposed to be mediated by a two-step fusion reaction, with the initial fusion event occurring with internal vesicles of endosomal transport intermediates, followed by a “back-fusion” reaction of viral nucleocapsid laden internal vesicles with the limiting membrane of late endosomes [6]. This mechanism has been disputed, and a more conventional fusion reaction directly with the limiting membrane of early endosomes proposed [2].

Membrane lipids play critical roles in the spatial organization of cells [7] and one feature of the internal vesicles of the endocytic pathway is that they contain a unique lipid, called either lysobisphosphatidic acid (LBPA) or bis(monoacylglycero)phosphate (BMP) [8, 9] LBPA is thought to promote membrane fusion and membrane invagination in a low pH-dependent manner [10, 11], and control the morphology and function of late endosomes and 4 multivesicular bodies. In this manuscript, we considered whether LBPA specifically promotes VSV G-mediated membrane fusion, and so provide a rationale for a two-step fusion reaction during VSV entry.

### ***Results and Discussion***

In order to determine the effect of LBPA on VSV G-mediated membrane fusion, we adapted a fluorescence resonance energy transfer (FRET)-based assay of lipid mixing [12]. In this assay, purified VSV particles are mixed with a liposome population consisting of liposomes labeled with both rhodamine-PE and NBD-PE and also unlabeled liposomes. A hemifusion or fusion event between a VSV particle and a labeled liposome results in a dilution of the FRET pairs and an increase in NBD

(donor) fluorescence, which is monitored by fluorescence spectroscopy. We produced liposomes that contained LBPA (20% molar ratio), and compared the extent of VSV-mediated fusion to liposomes without LBPA. A representative experiment is shown in Fig 2.1.

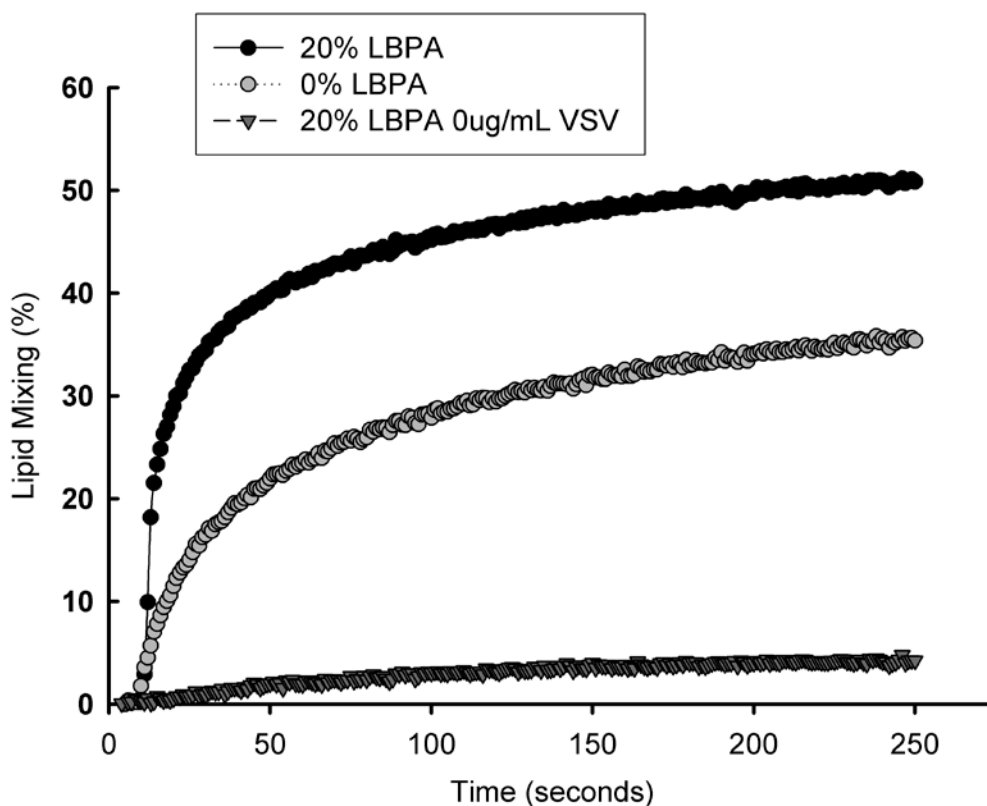


Figure 2.1 Lipid mixing of VSV with liposomes containing 4:5:1

POPC:POPS:cholesterol or 2:2:5:1 POPC:LBPA:POPS: cholesterol liposomes was initiated by a drop in pH, from 7.0 to 5.0. Lipid mixing was followed by monitoring fluorescence intensity at 530nm. 4:5:1 POPC:LBPA:POPS:cholesterol liposomes were also monitored following the pH drop in the absence of virus particles (-VSV).

Upon addition of VSV particles, both the initial rate of fusion and the extent of fusion were increased in the presence of LBPA. In the absence of virus particles, we

observed only limited fusion (Figure 2.1). The rate of VSV-induced lipid mixing within the first 10 seconds of the fusion reaction for liposomes with and without LBPA, at a range of pH values, is shown in Figure 2.2. At pH values below 6.5, liposomes with LBPA had significantly higher rates of lipid mixing than liposomes

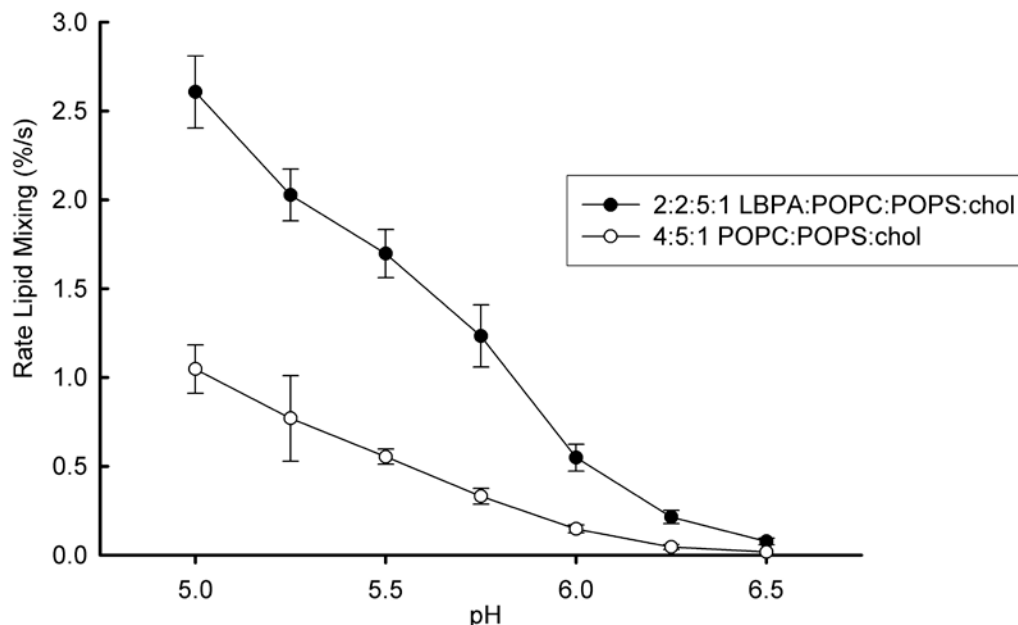


Figure 2.2 The rate of lipid mixing during the first 10 seconds of the reaction was determined over a range of pH conditions with labeled and unlabeled 4:5:1 POPC:POPS:cholesterol or 2:2:5:1 POPC:LBPA:POPS:cholesterol liposomes. Each data point is averaged from three individual assays and error bars represent standard deviation of the mean.

without LBPA.

The membranes of late endosomes have been reported to contain approximately 20% LBPA, which provided us a biologically relevant liposome composition to use in our initial experiments. In order to confirm the effect of LBPA on VSV G-mediated lipid mixing, we varied the amount of LBPA present in the



liposomes and examined the rate of lipid mixing. As shown in Figure 2.3, the rate of lipid mixing increases with increasing amounts of LBPA present in the liposome composition.

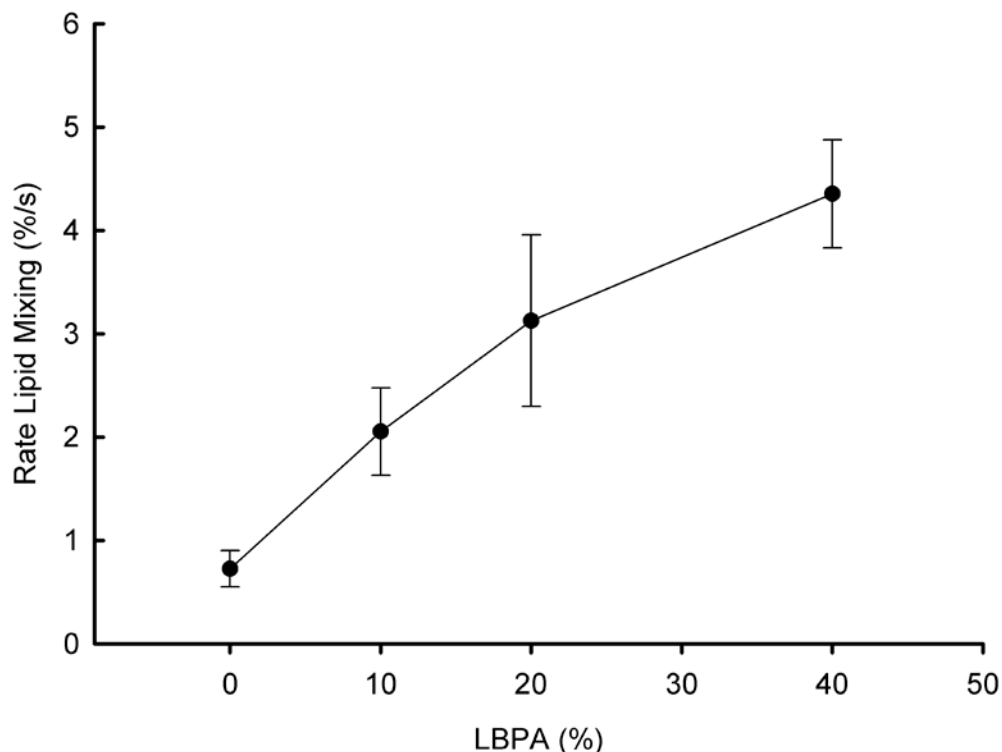


Figure 2.3 The rate of lipid mixing by VSV G increases with increased LBPA. The rate of lipid mixing during the first 10 seconds of the reaction, initiated by a change to pH 5, was examined with liposomes consisting of a range of LBPA quantities (from 0% to 40% of the liposome composition). The kinetics of lipid mixing was followed by monitoring NBD-PE fluorescence intensity at 530nm. Each data point is averaged from three individual assays and error bars represent standard deviation of the mean.

To date, a two-step fusion reaction has only been proposed during entry of VSV. We therefore examined whether LBPA might promote membrane fusion in other viral systems. We tested the effect of LBPA on membrane fusion mediated by

the hemagglutinin (HA) of influenza virus. As shown in Figure 2.4, our FRET-based assay clearly showed that when liposomes were fused with influenza virus (strain

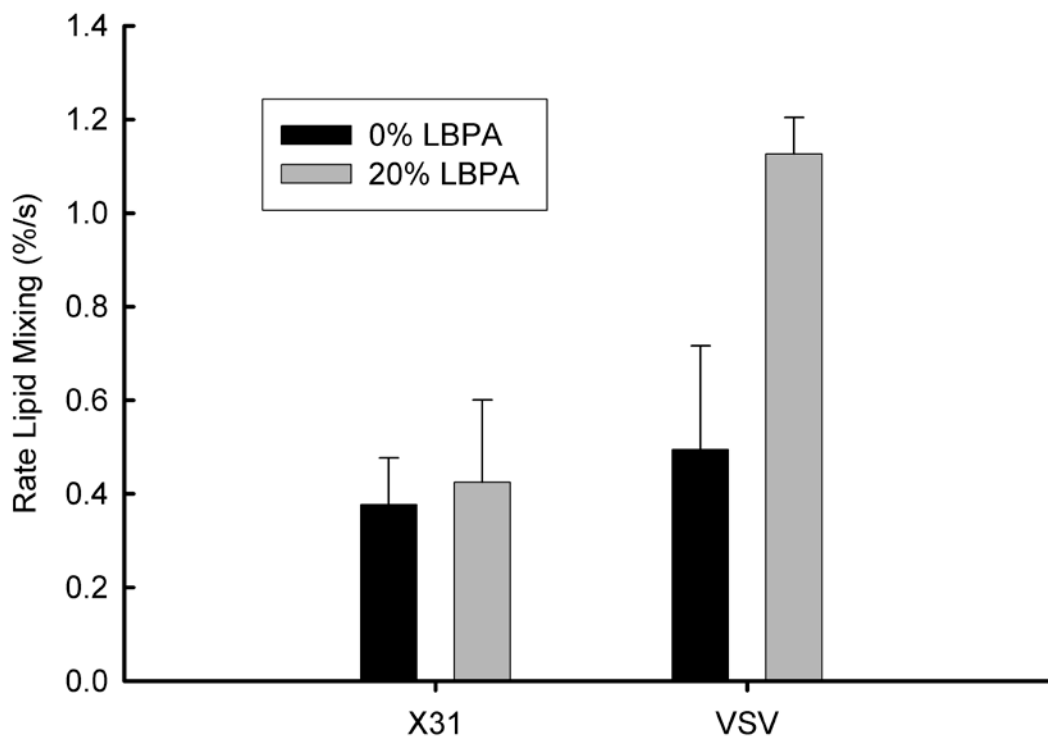


Figure 2.4 Virus specificity of lipid mixing rate acceleration by LBPA

A comparison of the rate of lipid mixing for VSV and influenza, both at a concentration of 100  $\mu\text{g/ml}$ , with 4:5:1 POPC:POPS:cholesterol or 2:2:5:1 POPC:LBPA:POPS:cholesterol liposomes at pH 5. Each bar is averaged from three individual assays and error bars represent standard deviation of the mean.

X:31), the addition of LBPA had no significant effect on the rate of fusion.

In contrast, LBPA had a profound effect on the rate of lipid mixing promoted by VSV G, suggesting that an interaction with LBPA is specific property of VSV.

Our data show that the endosome-specific phospholipid LBPA promotes membrane fusion mediated by the VSV G protein, and provide a rationale for a two-

step endosomal fusion reaction during VSV entry. While Gruenberg and colleagues considered that LBPA was involved in the back-fusion of VSV G-containing endosomal vesicles via the action of Alix [6], a direct role of LBPA in virus-induced fusion has not been addressed to date. It has previously been reported that the presence of LBPA in liposomes allows spontaneous fusion in a pH-dependent manner [10]. While some degree of LBPA-induced liposome fusion in the absence of virus particles was observed in our experiments, this was limited compared to fusion promoted by VSV G either in the presence or absence of LBPA. Our studies reveal an important role for LBPA in directly modulating VSV G mediated membrane fusion.

It has long been known that the VSV G protein can specifically interact with lipids, and a selective interaction with phosphatidylserine was originally proposed to account for virus-receptor interactions during entry into the cell [13], although this has recently been disputed [14]. In general, VSV G seems to have an affinity for negatively charged phospholipids [15, 16]. LBPA is also anionic, and so its charged head group could account for the enhancement of VSV G-mediated membrane fusion. However, the presence of negative charge does not appear to be the sole way of enhancing VSV G-mediated membrane fusion, as other negatively charged lipids such as palmitoyl-oleoyl-phosphatidic acid (POPA) showed only minimal enhancement of fusion (Figure 2.5). We assessed the VSV-G mediated fusion of liposomes containing other naturally occurring anionic lipids with similar chemical groups to LBPA including cardiolipin and POPG, which are commonly called polyglycerolipids. The FRET based assay showed that all polyglycerolipids enhanced the rate of VSV-G mediated lipid mixing beyond the anionic lipids POPS and POPA. Although VSV would not encounter POPG or cardiolipin during the normal course of host cell infection, this experiment shows that the chemical moieties common to this group of lipids could be involved in the mechanism of rate enhancement. Curiously, adding L-

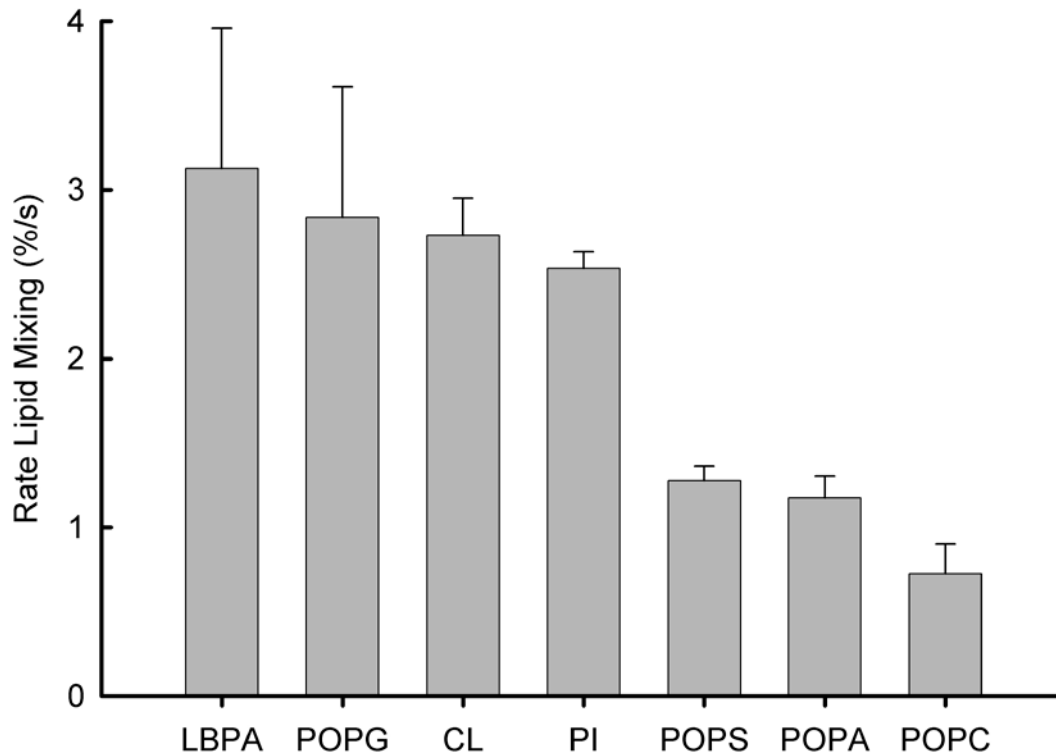


Figure 2.5 Effect of anionic lipids on the rate of lipid mixing by VSV-G

A comparison of the rates of lipid mixing for VSV with 2:2:5:1

POPC:X:POPS:cholesterol liposomes, where X is the anionic lipid indicated in the figure, at pH 5. Each bar is averaged from three individual assays and error bars represent standard deviation of the mean.

$\alpha$ -polyinositol (PI) showed a similar rate of lipid mixing enhancement as the polyglycerolipids. It is possible that the hydroxyl groups on the inositol ring can mimic the effect of the glycerol moiety on the polyglycerolipids. However, more experiments determining structure-function relationships would be required to validate this hypothesis.

One notable feature of our experiments is that whereas LBPA promoted VSV G-induced membrane fusion, it did not significantly affect influenza HA-mediated fusion. The fusion peptides of G and HA are very different; whereas influenza HA has

an external  $\alpha$ -helical fusion peptide that inserts into the lipid bilayer at an oblique angle [17, 18], VSV G has an internal fusion peptide composed of a bipartite loop structure that dips into the bilayer [3, 4]. The mechanism by which LBPA induces its differential effect of VSV is presently unclear, but it is likely to be related to an interaction with the viral fusion peptide. LBPA is a unique derivative of phosphatidylglycerol that shows unusual features for a membrane phospholipid [19]. One possibility for the effect of LBPA is related to its shape; however whether the lipid is cone, inverted cone or cylindrical is shape is unclear [19]. It has previously been reported that influenza virus HA-mediated fusion was inhibited by the inverted cone-shaped lipid lysophosphatidic acid and promoted by the cone-shaped oleic acid [20]. Due to the fact that LBPA showed no effect on influenza virus fusion (Fig 2.4), it is unlikely that an equivalent mechanism might underlie the effect of LBPA on VSV G—although such a conclusion would need more evidence for the physical behavior of LBPA in lipid bilayers.

There are two known serotypes of VSV, Indiana and New Jersey, which have approximately 50% identity in the G protein. VSV New Jersey is known to fuse at a lower pH than the Indiana serotype [21]. Due to this observation, we carried out the FRET based assay expecting that VSV New Jersey would show a greater enhancement of the lipid mixing rate upon the addition of LBPA because it would encounter more LBPA during the normal course of infection (Figure 2.6.) The result of the assay showed the exact opposite result. As shown in Figure 2.6, the lipid mixing rate for the Indiana strain with LBPA was significantly different (Students' t test,  $p \leq 0.05$ ), however, the lipid mixing rates for the New Jersey strain (Students' t test,  $p \leq 0.05$ ), however, the lipid mixing rates for the New Jersey strain were not significantly different ( $p \geq 0.05$ ). In order to compare the serotypes to each other at the pH values of 5 and 5.5, due to the quantification of virus by the Bradford assay.

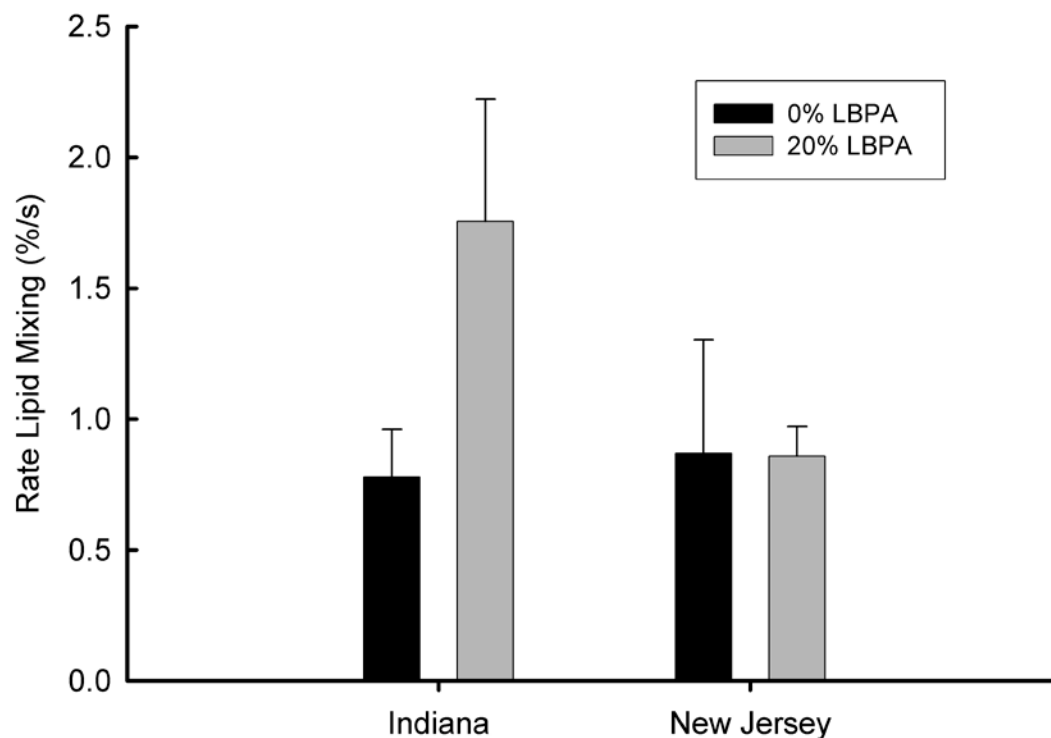


Figure 2.6 Comparison of lipid mixing for the two serotypes of VSV

A comparison of the rates of lipid mixing for VSV Indiana and New jersey with 4:5:1 POPC:POPS:cholesterol or 2:2:5:1 POPC:LBPA:POPS:cholesterol liposomes at pH 5. Each bar is averaged from three individual assays and error bars represent standard deviation of the mean.

If the ratio of lipid mixing with LBPA to lipid mixing without LBPA is not significantly different than one then there is no difference in the rate of lipid mixing. As shown in Figure 2.7, at pH 5 the ratio of lipid mixing for VSV Indiana was significantly different from 1 (One sample t test,  $p \leq 0.05$ ) but the ratio was not significantly different for the New Jersey serotype ( $p \geq 0.05$ ). The same trend was also evident for Indiana ( $p \leq 0.001$ ) and New Jersey ( $p \geq 0.05$ ) at pH 5.5. From this experiment it was concluded that one of the amino acids which is different between

the two serotypes could be responsible for the interaction with LBPA. In the fusion loops of VSV Indiana and New Jersey there is only a single amino acid difference. It seems unlikely that the change from an alanine residue to a glycine

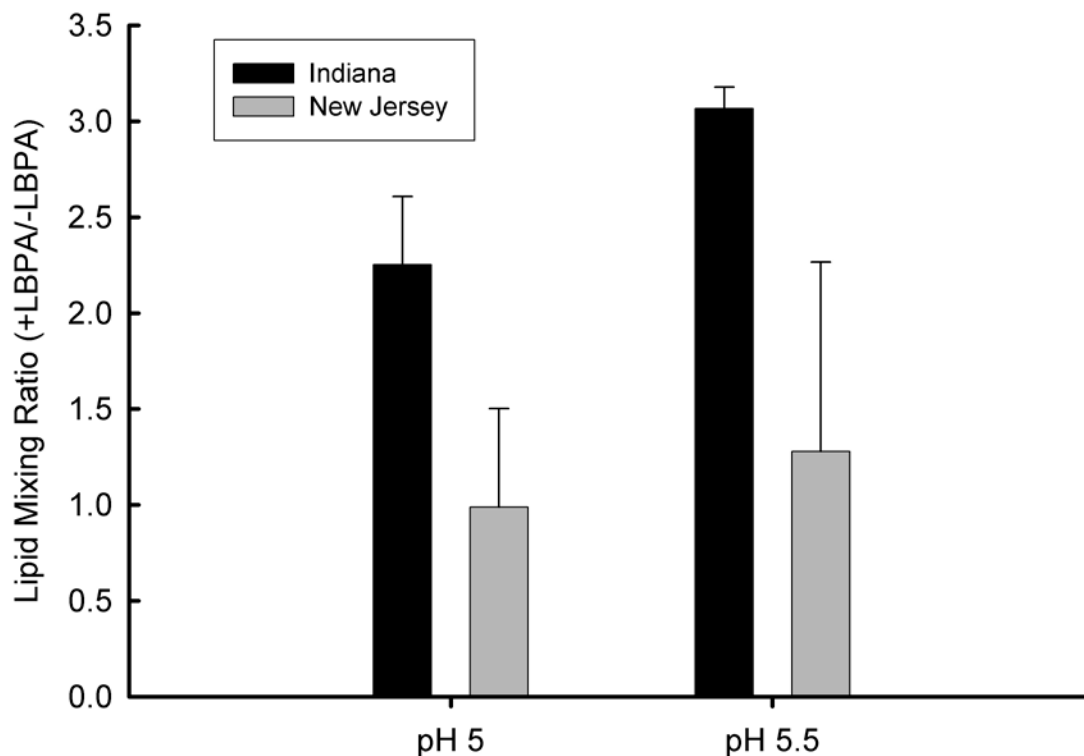


Figure 2.7 Comparison of the ratio of VSV-G mediated lipid mixing

A comparison of the rates of lipid mixing for VSV Indiana and New Jersey with 4:5:1 POPC:POPS:cholesterol or 2:2:5:1 POPC:LBPA:POPS:cholesterol liposomes at pH 5 or 5.5. Each bar was determined from three individual assays of each lipid composition and error bars represent propagated error.

residue could be responsible for such a dramatic effect; however an experiment examining the effect of mutating A117 to a glycine residue in VSV G Indiana and G117 to an alanine in VSV G New Jersey is clearly warranted.

The role of LBPA and the two-step model of VSV entry remains controversial.

It may be that depending on the composition of endosomal domains in different cell types, the virus can fuse either directly from the early endosome, or enter via back-fusion in late endosomes. As indicated in Fig. 2.3, cell types with higher amounts of fusion in late endosomes. As indicated in Fig. 2.3, cell types with higher amounts of endosomal LBPA may re-direct fusion toward the lumen of the MVB, rather than towards the cytosol. Such a variation in the site of fusion during virus entry might be facilitated by the relatively broad range of pH values (from approximately 6.2 to 5.0) that can activate the VSV G fusion machine.

VSV G is now considered to be a founder member of the so-called class III of viral fusion proteins [18]. Whereas the fusion loops of class III of viral fusion proteins, such as VSV G, show some similar features to class II fusion proteins, they retain distinctive structural properties [3]. Other class III viral fusion proteins include herpes simplex virus glycoprotein B and the baculovirus gp64 [22, 23]; it will be interesting to determine whether the promotion of VSV G mediated membrane fusion by LBPA is shared with these other viruses. It will also be interesting to address a direct role for LBPA in the context of other pathogens that navigate endosomal compartments during their entry into host cells [24].

### ***Materials and methods***

#### **Viruses and cells**

Vesicular stomatitis virus (VSV) Indiana, strain Orsay (ATCC) was propagated in BHK cells (ATCC). Briefly, 80-90% confluent roller bottles of BHK cells were infected at a MOI of 0.01. After 24h, the media was spun down and the supernatant was subjected to ultracentrifugation. The viral pellet was then applied to a sucrose step gradient to produce purified VSV. Viral protein concentration was determined using a Bradford protein assay. Influenza A virus, strain X:31 (H3N2), was grown in 11 day old embryonated eggs and allantoic fluid was collected to prepare concentrated virus stock as described for VSV.



## **Liposomes**

Palmitoyl-oleoyl-phosphocholine (POPC), palmitoyl-oleoyl-phosphoserine (POPS) lysobisphosphatidic acid (LBPA), and cholesterol were purchased from Avanti Polar Lipids (Alabaster, AL). Labeled phospholipids N-(7-nitrobenz-2-oxa-1,3-diazol-4-yl)-1,2-dihexadecanoyl-sn-glycero-3-phosphoethanolamine, triethylammonium salt (NBD-PE) and Lissamine<sup>TM</sup> rhodamine B 1,2-dihexadecanoyl-sn-glycero-3-phosphoethanolamine, triethylammonium salt (Rho-PE) were purchased from Invitrogen (Carlsbad, California). Large unilamellar vesicles (LUV) were prepared according to the extrusion method. Lipid films were obtained by subjecting chloroform dissolved lipid mixtures to high vacuum overnight. Lipid films were resuspended by addition of fusion buffer (20 mM MES, 30 mM Tris-HCl, pH 7) to 5 or 10 mM lipid concentration and incubated for 15 minutes at room temperature followed by vortexing for 15 minutes. Liposomes were then subjected to 10 freeze-thaw cycles, followed by 11 extrusions through 0.1  $\mu$ m polycarbonate membrane using an Avanti mini extruder. Liposomes labeled with the FRET pairs Rho-PE and NBD-PE were made in the same manner with the addition of 0.6% each of NBD-PE and Rho-PE to the chloroform-dissolved lipid mixture.

## **Lipid Mixing Assay**

Lipid mixing was determined using the method of Struck et al. (Struck et al., 1981). Unlabeled and labeled liposomes were mixed at a 4:1 ratio to a total concentration of 110  $\mu$ M lipid in fusion buffer at pH 7.0 with continuous stirring. Then VSV was added to a final concentration of 200  $\mu$ g/ml total viral protein unless otherwise noted. To initiate fusion, hydrochloric acid was added to the solution until the desired pH was achieved. To end the reaction and obtain a measurement of 100% lipid mixing, reduced Triton X-100 was added to a final concentration of 0.2%. Changes in fluorescence were measured using a QM-6SE spectrofluorimeter (Photon Technology

International, Birmingham, NJ) with excitation set at 467 nm and emission monitored at 530 and 581 nm. The extent of lipid mixing was determined using the formula:

$$F(\%) = \frac{f_t - f_0}{f_{100} - f_0} \times 100$$

where  $f_t$  is the fluorescence measurement at time  $t$ ,  $f_0$  is the initial fluorescence and  $f_{100}$  is the fluorescence after the addition of reduced Triton X-100. All measurements were taken in triplicate and averaged.

## REFERENCES

1. Marsh, M. and A. Helenius, *Virus entry: Open sesame*. Cell, 2006. **124**(4): p. 729-740.
2. Johannsdottir, H.K., R. Mancini, J. Kartenbeck, L. Amato, and A. Helenius, *Host cell factors and functions involved in vesicular stomatitis virus entry*. Journal of Virology, 2008. **published ahead of print on 29 October 2008**.
3. Sun, X., S. Belouzard, and G.R. Whittaker, *Molecular architecture of the bipartite fusion loops of vesicular stomatitis virus glycoprotein G, a class III viral fusion protein*. Journal of Biological Chemistry, 2008. **283**(10): p. 6418-6427.
4. Roche, S., et al., *Structures of vesicular stomatitis virus glycoprotein: membrane fusion revisited*. Cellular and Molecular Life Sciences, 2008. **65**(11): p. 1716-1728.
5. Gaudin, Y., *Reversibility in fusion protein conformational changes. The intriguing case of rhabdovirus-induced membrane fusion*. Subcellular Biochemistry, 2000. **34**: p. 379-408.
6. Le Blanc, I., et al., *Endosome-to-cytosol transport of viral nucleocapsids*. Nature Cell Biology, 2005. **7**(7): p. 653-U25.
7. van Meer, G., D.R. Voelker, and G.W. Feigenson, *Membrane lipids: where they are and how they behave*. Nature Reviews Molecular Cell Biology, 2008. **9**(2): p. 112-124.
8. Kobayashi, T., et al., *Localization of lysobisphosphatidic acid-rich membrane domains in late endosomes*. Biological Chemistry, 2001. **382**(3): p. 483-485.
9. Piper, R.C. and D.J. Katzmann, *Biogenesis and function of multivesicular bodies*. Annual Review of Cell and Developmental Biology, 2007. **23**: p. 519-547.

10. Kobayashi, T., et al., *Separation and characterization of late endosomal membrane domains*. Journal of Biological Chemistry, 2002. **277**(35): p. 32157-32164.
11. Matsuo, H., et al., *Role of LBPA and Alix in multivesicular liposome formation and endosome organization*. Science, 2004. **303**(5657): p. 531-534.
12. Struck, D.K., D. Hoekstra, and R.E. Pagano, *Use of resonance energy-transfer to monitor membrane-fusion*. Biochemistry, 1981. **20**(14): p. 4093-4099.
13. Schlegel, R., et al., *Inhibition of VSV binding and infectivity by phosphatidylserine - is phosphatidylserine a VSV-binding site?* Cell, 1983. **32**(2): p. 639-646.
14. Coil, D.A. and A.D. Miller, *Phosphatidylserine is not the cell surface receptor for vesicular stomatitis virus*. Journal of Virology, 2004. **78**(20): p. 10920-10926.
15. Carneiro, F.A., et al., *Probing the interaction between vesicular stomatitis virus and phosphatidylserine*. European Biophysics Journal with Biophysics Letters, 2006. **35**(2): p. 145-154.
16. Eidelman, O., et al., *pH-dependent fusion induced by vesicular stomatitis-virus glycoprotein reconstituted into phospholipid-vesicles*. Journal of Biological Chemistry, 1984. **259**(7): p. 4622-4628.
17. Han, X., et al., *Membrane structure and fusion-triggering conformational change of the fusion domain from influenza hemagglutinin*. Nature Structural Biology, 2001. **8**(8): p. 715-720.
18. White, J.M., et al., *Structures and mechanisms of viral membrane fusion proteins: Multiple variations on a common theme*. Critical Reviews in Biochemistry and Molecular Biology, 2008. **43**(3): p. 189-219.
19. Holopainen, J.M., et al., *Intermolecular interactions of lysobisphosphatidic*

- acid with phosphatidylcholine in mixed bilayers*. Chemistry and Physics of Lipids, 2005. **133**(1): p. 51-67.
20. Chernomordik, L.V., et al., *An early stage of membrane fusion mediated by the low pH conformation of influenza hemagglutinin depends upon membrane lipids*. Journal of Cell Biology, 1997. **136**(1): p. 81-93.
  21. Martinez, I. and G.W. Wertz, *Biological differences between vesicular stomatitis virus Indiana and New Jersey serotype glycoproteins: Identification of amino acid residues modulating pH-dependent infectivity*. Journal of Virology, 2005. **79**(6): p. 3578-3585.
  22. Heldwein, E.E., et al., *Crystal structure of glycoprotein B from herpes simplex virus 1*. Science, 2006. **313**(5784): p. 217-220.
  23. Kadlec, J., et al., *The postfusion structure of baculovirus gp64 supports a unified view of viral fusion machines*. Nature Structural & Molecular Biology, 2008. **15**(10): p. 1024-1030.
  24. Gruenberg, J. and F.G. van der Goot, *Mechanisms of pathogen entry through the endosomal compartments*. Nature Reviews Molecular Cell Biology, 2006. **7**(7): p. 495-504.

## CHAPTER 3

### LIPID MIXING PROMOTED BY NOVEL CORONAVIRUS FUSION PEPTIDES

#### *Introduction*

In order to produce progeny a virus must gain entry into a host cell, replicate its genetic information and component proteins, and assemble and exit the cell. In order to aid in its entrance into the cell, viruses have developed surface proteins often called fusion proteins. These proteins are able to disrupt the membrane of a host cell and facilitate the delivery of the virus's nuclear material into the cell's cytoplasm. The part of the fusion protein that actually inserts into the outer leaflet of the cell's plasma membrane is termed the fusion peptide.

The fusion peptide has been defined as having several properties [1]. First, a fusion peptide must be triggered by environmental conditions to be exposed from a protected area within the fusion protein. It is logical that a fusion peptide would be protected during the majority of the virus's life cycle to prevent non-productive fusion events. Several conditions have been proposed to cause fusion peptide exposure [2] including low pH, receptor binding, a combination of low pH and receptor binding and proteolytic cleavage. Secondly, a fusion peptide must insert into the hosts cell's membrane, whether at the cell surface or inside an endosome. Finally, the fusion peptide must be involved in the formation of the fusion pore possibly with other parts of the fusion protein such as the transmembrane region. Although it seems that all of these stipulations are self-evident, conclusively proving that a specific peptide has all of these properties is a challenging endeavor.

Several class I fusion peptides have been identified, and are listed in Table 3.1 [3]. Although fusion peptides are not conserved among viruses, several characteristics have proven similar among known class I fusion protein fusion peptides[4]. Despite the fact that between families, fusion peptides are not conserved,

Table 3.1 Amino acid sequences of fusion peptides from Class I fusion proteins. Vertical arrows represent protease cleavage sites. Furin/PC refers to cleavage by a protease in the proprotein convertase family.

| <i>External (N-terminal)</i> |                             | <i>Cleaving Protease</i> |
|------------------------------|-----------------------------|--------------------------|
| Influenza HA:                | ↓GLFGAIAGFIENGWEG           | Trypsin/Furin            |
| Sendai F1:                   | ↓FFGAVIGTIALGVATA           | Trypsin                  |
| RSV F1:                      | ↓FLGFLLGVGSAIASGV           | Furin/PC                 |
| HIV gp41:                    | ↓AAIGALFLFGLGAAGSTMGAA      | Furin/PC                 |
| LASV GP-2:                   | ↓GTFTWTLSDSGGKDTPGGYCL      | SKI-1/S1P                |
| JUNV GP-2:                   | ↓AFFSWSLTDSSGKDTPGGYCIE     | SKI-1/S1P                |
| <br><i>Internal</i>          |                             |                          |
| Ebola GP:                    | ↓..22aa..GAAIGLAWIPYFGPAA   | Furin/PC                 |
| ASLV gp37:                   | ↓..22aa..IFASILAPGVAAAQAL   | Furin/PC                 |
| Nipah F:                     | ↓..??aa..LAGVIMAGVAIGIATAAQ | Cathepsin                |
| Hendra F:                    | ↓..??aa..LAGVVMAGIAIGIATAAQ | Cathepsin                |

within a virus family they are one of the most highly conserved protein sequences. Usually that sequence is rich in alanines and glycines and can contain several of the bulky hydrophobic amino acids. Overall, the peptide typically has an intermediate level of hydrophobicity. An external fusion peptide is often considered to terminate at the first positively charged residue while an internal fusion peptide usually has a positively charged residue on the N-terminus and either a positively or negatively charged amino acid on the C-terminus. The total length of an external fusion peptide is usually 20-36 amino acids while internal fusion peptides are usually slightly shorter 16-20 amino acids. From the amino acid sequence of a fusion peptide, one should be able to predict a high potential to bind membranes. Additionally, it has been found that fusion peptides tend to insert themselves into membranes at an oblique angle. Due to the fact that a peptide can possess all of these characteristics and still not be a viral fusion peptide, experimental evidence must be collected to verify the identity of the peptide.

## **Fusion peptide identification and characterization**

Several techniques have been used to identify fusion peptides in viral fusion proteins with various degrees of success. Even though the characteristics of a fusion peptide listed previously can help identify the peptide, the peptide must still be shown to be exposed by an environmental trigger, insert into the host cell's membrane, and promote formation of the fusion pore as previously described. Whereas a single experiment cannot prove a peptide is able to perform all of these tasks, usually several experiments can provide enough information to determine the identity of a fusion protein's fusion peptide. One of the most helpful tools used in identifying fusion peptides is X-ray crystallography. By obtaining a crystal structure of a fusion protein in its pre- and post-fusion forms, the fusion peptide can usually be identified. Crystal structures proved to be of great assistance in determining the identity of the fusion peptide of the influenza HA protein[5], and more recently, the fusion loops of the vesicular stomatitis virus G protein[6, 7]. Unfortunately, fusion protein crystal structures cannot be used alone to determine the identity of the fusion peptides. Often times, it is extremely difficult to crystallize a virus fusion protein due to their usually extensive glycosylation patterns and the difficulty of obtaining enough protein to even attempt to grow a crystal. Additionally, a crystal structure cannot prove that a peptide is inserted into a membrane or that it forms the fusion pore. Fortunately, other methods are available that can answer those questions. Another widely used method to identify a fusion peptide is sequence and mutational analysis. Although the fact that a protein sequence is highly conserved among a virus family does not prove it is the fusion peptide, it can strengthen the argument when taken into consideration with other experimental data [2]. Mutational analysis commonly consists of replacing the residues of the fusion peptide, usually with alanine, and determining the effect on fusion. When critical residues of the fusion peptide are significantly changed, fusion



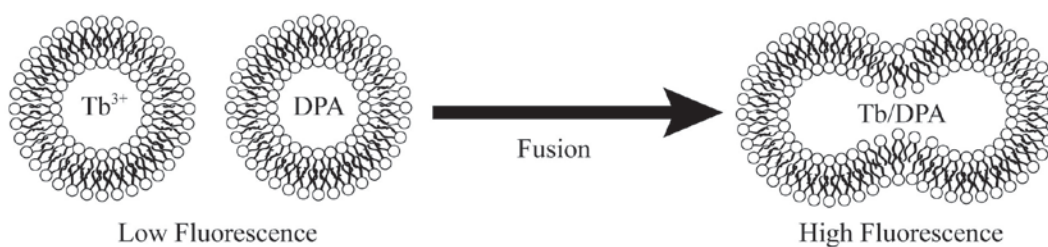
is usually decreased as determined by cell-cell fusion assay, pseudoviral or recombinant virus infection assay. The identification of residues that have a large influence on fusion does not guarantee that they are located in the fusion peptide, but it is a common way to corroborate evidence from subsequent experiments. The use of spectroscopic techniques such as circular dichroism (CD) and Fourier transform infrared spectroscopy (FTIR) have also been used to probe fusion peptides. Many fusion peptides have high degrees of alpha helical character that can be identified by these methods. Additionally, a variation of FTIR using attenuated total reflectance (ATR-FTIR) can be used to determine the angle of insertion of a peptide into a membrane bilayer. Characteristically, fusion peptides insert into lipid bilayers at an oblique angle[8] from approximately 20-60° in other words, at an angle more parallel to the bilayer than perpendicular. However, again, these spectroscopic methods cannot determine the fusion peptide of a fusion protein since they only characterize its structure. There have been several literature reports of using photolabeling and cross linking to a membrane to identify a fusion peptide. This method, though technically challenging, can be used to establish that the sequence in question is environmentally triggered and inserts itself into membranes. Although the results can be difficult to interpret at times[9], it has been useful in identifying fusion peptides from several viruses[10]. Finally, various assays with liposomes have been used to probe the mechanism of fusion by viruses, fusion proteins, and fusion peptides[11]. This assay, when carried out with a synthetic peptide of the proposed fusion peptide, can provide support that the peptide can insert into membranes and cause lipid mixing and/or fusion pore formation.

### **Liposome based fusion assays**

Assays with liposomes and fusion peptides can be used to determine whether the fusion peptide being examined has the ability to insert into the lipid bilayer and

promote lipid mixing. Steps which must be accomplished in order to advance towards the formation of the fusion pore. In vitro liposome experiments require the use of a synthetic peptide to approximate the behavior of the viral fusion protein. Although one can imagine that the conformation and behavior of a peptide in the context of a protein could be vastly different than as an isolated peptide, experiments have shown that these peptides are sensitive to mutations and environmental factors (such as pH) as predicted by experiments with the entire fusion protein or virus[12]. There are two main types of fusion assays involving liposomes that can be used to assess the fusion abilities of a peptide, as shown in Figure 3.1 [13].

A.



B.

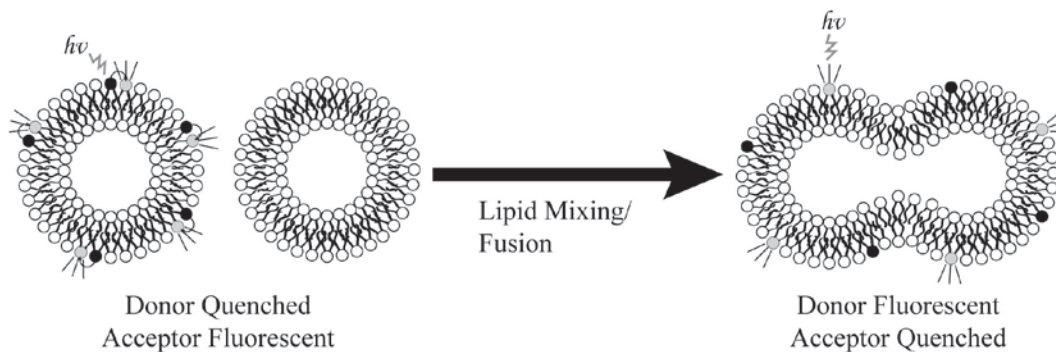


Figure 3.1 Lipid mixing/fusion assays involving liposomes. A. Terbium-dipicolinic acid (DPA) content mixing assay. B. FRET based dye dilution lipid mixing assay.

The first is content mixing assays which involve the inclusion of a reporter molecule

in the liposomes. When peptide induced leakage of the liposome contents occurs, the fluorescence of the reporter molecule changes indicating a perturbation of the lipid bilayer. Commonly, the interaction between terbium (Tb) and dipicolinic acid (DPA) or aminonaphthalene trisulfonic acid (ANTS) and bis(pyridinium) bromide (DPX) are used [13]. The second type of fusion assay involving liposomes, which was used to explore the activities of coronavirus fusion peptides in this report, is referred to as dye dilution assays. The most commonly used assay of this type was developed by Struck et. al[14] and has been used to examine the properties of several fusion peptides to date[12]. The most frequently used dyes for a dye dilution assay are 7-nitrobenz-2-oxa-1,3-diazol-4-yl (NBD) and rhodamine (Rh) which can be tethered to a lipid such as phosphatidylethanolamine (PE) or cholesterol, but other appropriate FRET pairs can also be utilized[15]. The energy transfer is a non-radiative process dependent on the extent of overlap of the emission spectrum of the donor with the absorption spectrum of the acceptor, orientation of the dye molecules, and distance[16]. This assay takes advantage of the effect of distance on the fluorescence of the donor and acceptor. In the assay, two populations of liposomes are present, one labeled with both NBD-PE and rhodamine-PE and one without either dye. When the fusion peptide is added and the pH lowered, the fusion peptide induces aggregation and fusion. When a labeled liposome fuses with an unlabeled liposome, dye dilution takes place, increasing the distance between the FRET pairs and decreasing energy transfer. The result is a measureable increase in the fluorescence intensity of the donor dye and a decrease in the intensity of the acceptor dye. To determine full fusion, at the conclusion of the reaction, reduced Triton-X100 is added to the reaction. The total extent of fusion promoted by the peptide can then be calculated with the following formula[17]:

$$F_t(\%) = [(I_t - I_0)/(I_t - I_0)] \times 100$$

The dye dilution assay is generally more sensitive than content mixing assays, and require less fluorescent probe. The procedure involved in carrying out the lipid mixing assay using the dye dilution method will be discussed in detail in the materials and methods section below. The liposome FRET assay has been used to probe the fusion peptides of a variety of viruses including influenza [18, 19], HIV [20, 21], Sendai virus [22], Ebola [23, 24] and severe acute respiratory syndrome-associated coronavirus (SARS) [25, 26].

### **Coronaviruses**

Coronaviruses are large, enveloped, positive-stranded RNA viruses known to cause disease in birds and mammals, including humans [27]. Of special interest to this report are infectious bronchitis virus (IBV), feline infectious peritonitis virus (FIPV), feline enteric coronavirus (FECV) and severe acute respiratory syndrome virus (SARS). Although these viruses belong to different coronavirus antigenic groups, and infect different species, they show structural similarities among their spike proteins. The coronavirus spike protein is generally said to contain two domains. The S1 domain contains the receptor binding site which is significantly different among these viruses. For example, it is known that FIPV uses feline aminopeptidase N (APN) as a cellular receptor, while SARS is known to use angiotensin-converting enzyme 2 (ACE2) [28]. Additionally, it has been recently shown that both of these viruses can also use C-type lectins to enhance attachment to cells [29, 30]. The S2 domain is thought to be responsible for fusion with the host cell membrane. Currently, the identity of the fusion peptide in all of these viruses is unknown. There have been several reports in the literature of segments of the SARS S protein that has an influence on fusion [25, 26, 31, 32] and could be the fusion peptide. The authors of

these papers suggest three sections or peptides within the SARS S protein that potentially have a role in fusion. They were identified using the Wimley and White hydrophobicity interface scale which rated these S protein regions as being highly likely to associate with lipid membranes. The first region, termed WWI is N-terminal to our proposed fusion peptide and consists of residues 770-788. The second region WWII is C-terminal to our proposed fusion peptide and consists of residues 864-886. Both of these fusion peptides are N-terminal to the two heptad repeat regions. More recently the authors proposed a third region near the transmembrane region which they named SARS<sub>PTM</sub> which they identify as consisting of residues 1187-1200. The authors carry out numerous biochemical assays with synthetic peptides including peptide-lipid membrane binding assays, liposome based lipid mixing and content leakage assays, circular dichroism spectroscopy and others in order to show that the proposed fusion peptides can bind and penetrate membranes and promote fusion. Additionally, they show that these regions have other properties, such as structure, which are similar to known fusion peptides. A second group has mutated the region identified as WWI and shown that mutations in this region can reduce fusion in cell-cell fusion assays as much as 70%. Despite their efforts, the data presented in the literature thus far supporting these three fusion peptides is inconclusive. Previous work in the Whittaker lab has revealed that the coronavirus IBV Beaudette strain contains a furin cleavage site between residues 687 and 690, named S2'[33]. Upon exposure of the IBV Beaudette spike protein to furin, a cleavage product can clearly be observed. Alignment of IBV Beaudette with other coronaviruses shows that an arginine residue, which is a potential protease cleavage site, and the sequence C-terminal to the cleavage site is highly conserved among coronaviruses, Figure 3.2.

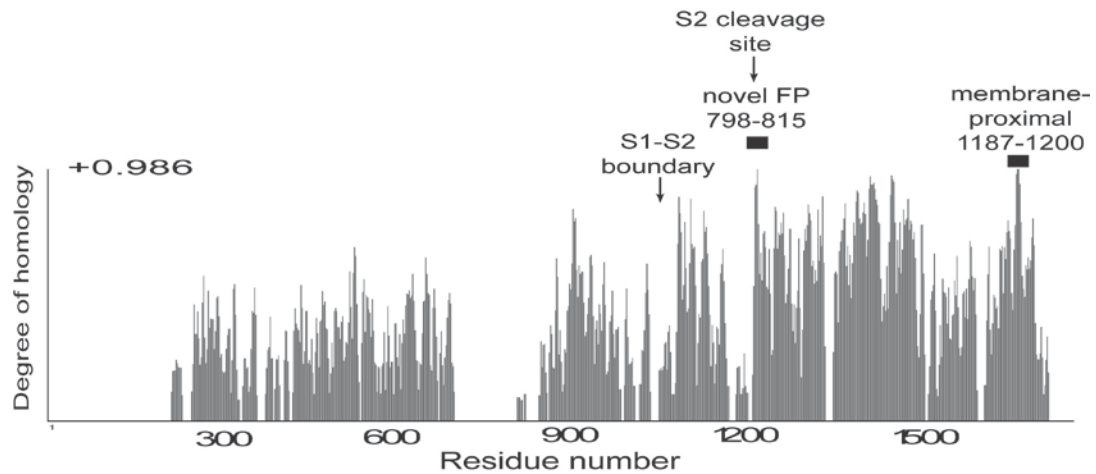


Figure 3.2 Alignment of coronavirus spike proteins. The graph shows the degree of homology in the spike protein of the coronavirus family. Shown below are the sequences of the proposed fusion peptides in the three classes of coronaviruses. (Madu, Roth, et. al, *J. Virol.*, 2009).

These observations lead to the examination of this cleavage site in SARS and the prediction that cleavage at this location exposed the fusion peptide. It has been determined that mutations at the S2' cleavage site in SARS inhibits fusion. Additionally, replacing the two basic residues K796 and R797 with a furin cleavage site enhances cleavage and fusion[33]. Subsequently, it has been shown that mutation to the region succeeding the cleavage site has a deleterious effect on fusion in cell-cell fusion assay and infection when the mutations are incorporated into pseudoviral particles [34].

Closer examination of the coronaviruses shows that one member of the family, FIPV, contains a mutation at the otherwise conserved protease cleavage site, as shown in Table 3.2.

Table 3.2 Sequence alignment of coronavirus S2' cleavage sites. Vertical bars represent potential protease cleavage sites.

|                   |                        |
|-------------------|------------------------|
| SARS CoV (Urbani) | PTKR SFIEDLLFNKVTLAD   |
| IBV (Bdtte)       | RRRR SVIEDLLFTSVESVG   |
| FIPV (79-1146)    | RK Y GSAIEDLLFDKVVTS G |
| FECV (79-1683)    | RKYR SAIEDLLFDKVVTS G  |

Due to the fact that FIPV has a glycine instead of arginine residue at the S2' site, it is reasonable to assume that the cleavage site that exposes the fusion peptide could have shifted in this case. After analyzing the cleavage site preference of cathepsin L and B which are known to cleave the feline coronavirus S proteins, we determined that it is possible that the FIPV spike protein could be cleaved one or two residues N-terminal to the fusion peptide, as indicated in Figure 3.3 by the vertical lines. Addition of one or two residues to the N-terminus of the FIPV fusion peptide could have a profound effect on its ability to mediate lipid mixing.

Based on all of these observations we decided to investigate the synthetic fusion peptides of the coronaviruses IBV, SARS, FIPV and FECV to determine whether these peptides are capable of associating with membranes and promoting lipid mixing. Although these experiments do not prove that these sequences are indeed the fusion peptides of these viruses, they lend support to the hypothesis that these regions are capable of insertion into the host cell membrane and that they could be responsible for the formation of the fusion pore.

## ***Materials and Methods***

### **Peptides**

The coronavirus fusion peptides, Table 1, were synthesized using solid phase techniques by New England Peptide (Gardner, MA).

Table 3.3 Synthesized coronavirus fusion peptide sequences.

| Coronavirus      | Fusion Peptide Sequence      |
|------------------|------------------------------|
| IBV              | SLIEDLLFTSVESGCGKKK          |
| SARS             | SFIEDLLFNKVTADAGFMKQYGCGKKKK |
| SARS short       | SFIEDLLFGCGKKKK              |
| SARS LLF-AAA     | SFIEDAAAGCGKKK               |
| FIPV             | YGSAIEDLLFDKVVTSGLGGGCGKKKK  |
| FIPV-Y           | GSAIEDLLFDKVVTSGLGGGCGKKKK   |
| FECV             | SAIEDLLFDKVVTSGLGGGCGKKKK    |
| Negative Control | GCGKKKK                      |

The GCGKKKK linker was included to promote liposome association as described for the host-guest fusion peptide system of Han and Tamm [18]. Purity as determined by HPLC was greater than 95%, and mass identification was performed by MALDI-TOF mass spectrometry. Peptides were resuspended to 5 mg/ml in sterile MilliQ water. Control peptides were treated identically and had the sequence SIRYSFCGNGRHV for circular dichroism spectroscopy experiments and GCGKKKK for lipid mixing experiments.

### **Liposomes**

1-Palmitoyl-2-Oleoyl-*sn*-Glycero-3-Phosphocholine (POPC), 1-Palmitoyl-2-Oleoyl-*sn*-Glycero-3-[Phospho-L-Serine] (Sodium Salt) (POPS), 1-Palmitoyl-2-Oleoyl-*sn*-Glycero-3-[Phospho-*rac*-(1-glycerol)] (Sodium Salt) (POPG), and cholesterol were



purchased from Avanti Polar Lipids (Alabaster, AL). Labeled phospholipids *N*-(7-nitrobenz-2-oxa-1,3-diazol-4-yl)-1,2-dihexadecanoyl-*sn*-glycero-3-phosphoethanolamine, triethylammonium salt (NBD-PE) and Lissamine<sup>TM</sup> rhodamine B 1,2-dihexadecanoyl-*sn*-glycero-3-phosphoethanolamine, triethylammonium salt (Rho-PE) were purchased from Invitrogen (Carlsbad, California). Large unilamellar vesicles (LUV) were prepared according to the extrusion method. Lipid films were obtained by subjecting chloroform dissolved lipid mixtures of 1:3:1 POPC:POPS:cholesterol in the case of IBV, SARS and SARS short, and 4:1 POPG:cholesterol in the case of FIPV, FIPV-Y, and FECV, to high vacuum overnight. Lipid films were resuspended by addition of fusion buffer (5 mM HEPES, 5 mM MES, 5mM sodium succinate, 150 mM sodium chloride, pH 7) to 5 or 10 mM lipid concentration and incubated for 15 minutes at room temperature followed by vortexing for 15 minutes. Liposomes were then subjected to 10 freeze-thaw cycles, followed by 11 cycles of extrusion through 0.1  $\mu$ m polycarbonate membrane using an Avanti mini extruder. Liposomes labeled with the FRET pairs Rho-PE and NBD-PE were made in the same manner with the addition of 0.6% each of NBD-PE and Rho-PE to the chloroform-dissolved lipid mixture.

### **Lipid Mixing Assay**

Lipid mixing was determined using the method of Struck et al. [14]. Unlabeled and labeled liposomes were mixed at a 4:1 ratio to a total concentration of 110  $\mu$ M lipid in fusion buffer at pH 7.0. Hydrochloric acid was added to the solution until the desired pH was achieved. To initiate fusion, peptide was added to the specified concentration. To end the reaction and obtain a measurement of 100% lipid mixing, reduced Triton X-100 was added to a final concentration of 0.2%. Changes in fluorescence were measured using a QM-6SE spectrofluorimeter (Photon Technology International,

Birmingham, NJ) with excitation set at 467 nm and emission monitored at 530 and 581 nm. The extent of lipid mixing was determined using the previously discussed formula. All measurements were taken in triplicate and averaged.

## ***Results***

### **IBV**

In order to determine the capability of the proposed coronavirus fusion peptide to promote lipid mixing, the IBV fusion peptide shown in Table 3.1 was subjected to the described lipid mixing assay. The extent of lipid mixing was monitored as an increase in NBD fluorescence over three trials and the averages at different peptide concentrations is represented in Figure 3.3. The IBV fusion peptide was able to promote a high amount of lipid mixing at pH 5, and was almost completely unable to promote lipid mixing at pH 7.

### **SARS**

To determine whether the ability of the proposed fusion peptide to promote lipid mixing was characteristic of this sequence in the other members of the coronaviridae family, we examined the SARS fusion peptide under similar conditions to that of the IBV fusion peptide. The extent of lipid mixing promoted by the SARS fusion peptide and monitored in the FRET-dilution assay are shown in Figure 3.4. From these experiments it is evident that the SARS fusion peptide is able to mediate lipid mixing at both pH 5 and 7, unlike the IBV peptide which is only able to mediate lipid mixing at pH 5. Although it is possible for the SARS fusion peptide to mediate lipid mixing in both pH conditions, the SARS fusion peptide is able to promote a higher extent of lipid mixing at lower peptide concentrations at pH 5. It should be noted that the negative control peptide GCGKKK, which increases the solubility of the fusion peptide, was included in these lipid mixing experiments to show that this region does

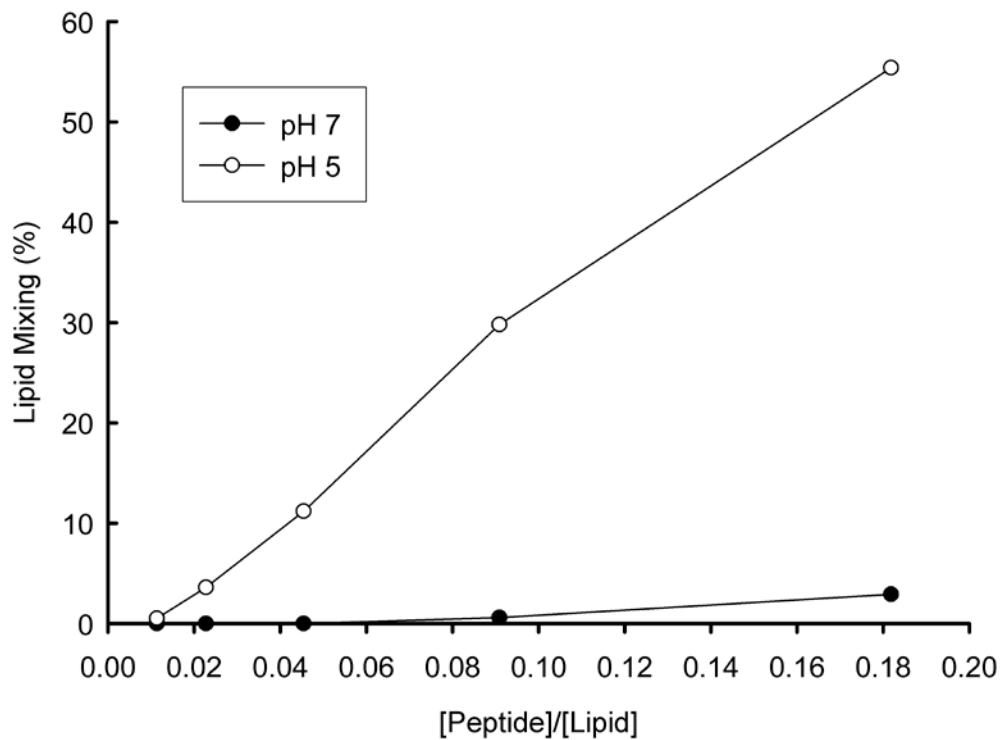


Figure 3.3 The IBV S2 fusion peptide induces lipid mixing in liposomes Extent of lipid mixing was determined by varying the ratios of IBV S2 fusion peptide with total concentration of labeled and unlabeled 1:3:1 POPC:POPS:cholesterol liposomes at pH 5 and pH 7. The kinetics of mixing was followed by monitoring fluorescence intensity at 530nm upon addition of peptide.

not promote lipid mixing independent of the fusion peptide.

To further examine the pH dependence of the SARS fusion peptide, we determined its ability to mediate lipid mixing at pH values in between pH 5 and 7. It can be seen in Figure 3.5 that lipid mixing mediated by the SARS fusion peptide has an almost linear dependence on pH.

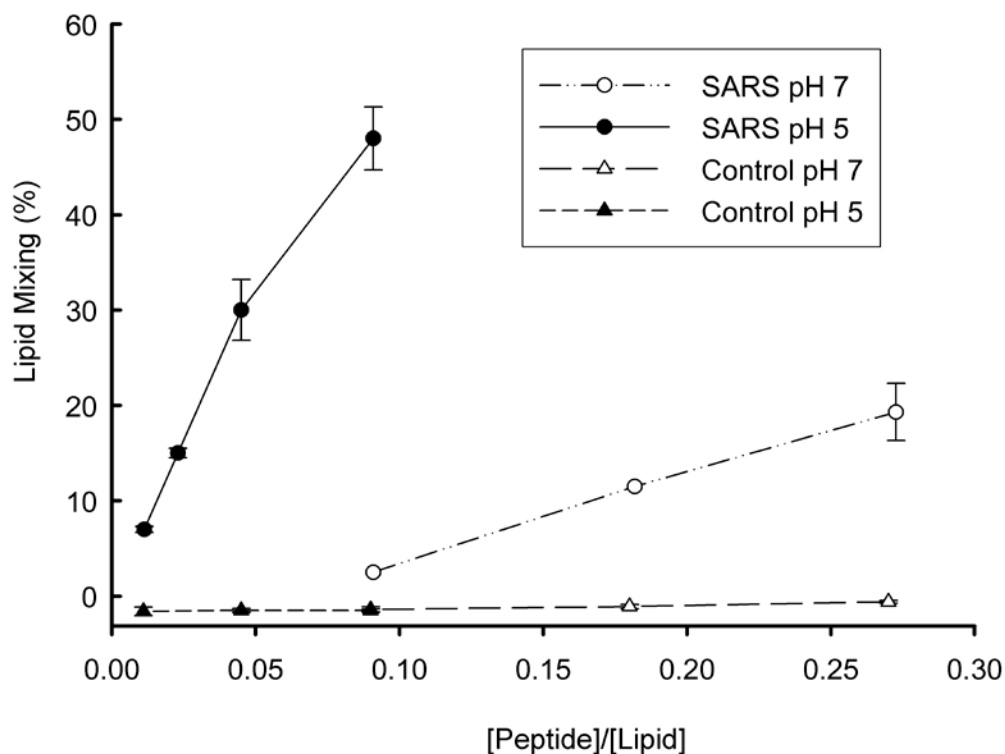


Figure 3.4 The SARS-CoV S2 fusion peptide induces lipid mixing in liposomes.

Extent of lipid mixing was determined by varying the ratios of SARS-CoV S2 fusion peptide or control peptide concentrations with total concentration of labeled and unlabeled 1:3:1 POPC:POPS:cholesterol liposomes at pH 5 and pH 7. The kinetics of mixing was followed by monitoring fluorescence intensity at 530nm upon addition of peptide. Each data point is averaged from three individual assays and error bars represent standard deviation of the mean.

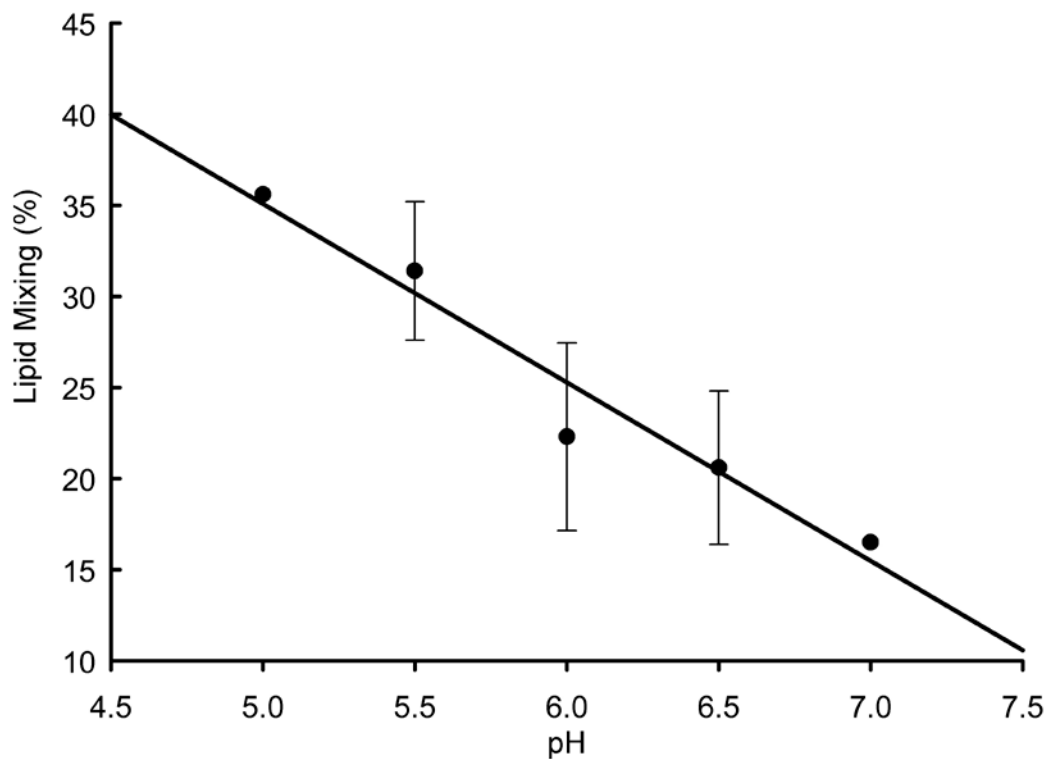


Figure 3.5 Effect of pH on SARS-CoV S2 fusion peptide mediated lipid mixing. The kinetics of lipid mixing was followed by monitoring NBD-PE fluorescence intensity at 530nm in varying pH environments (from pH 7.0 to pH 5.0). Each data point is averaged from three individual assays and error bars represent standard deviation of the mean.

A significant experiment which can lend support to the identification of a fusion peptide is to show that mutations that abolish fusion in the cell based assays also prevent lipid mixing in liposome based assays with the isolated, synthetic peptide. Figures 3.6 and 3.7 show the effect of mutating key residues in the SARS fusion peptide, LLF, which were identified in cell-cell fusion and pseudovirus assays [34]. Mutation of these residues to alanine completely destroys the ability of the peptide to promote lipid mixing at both pH 5 and 7.

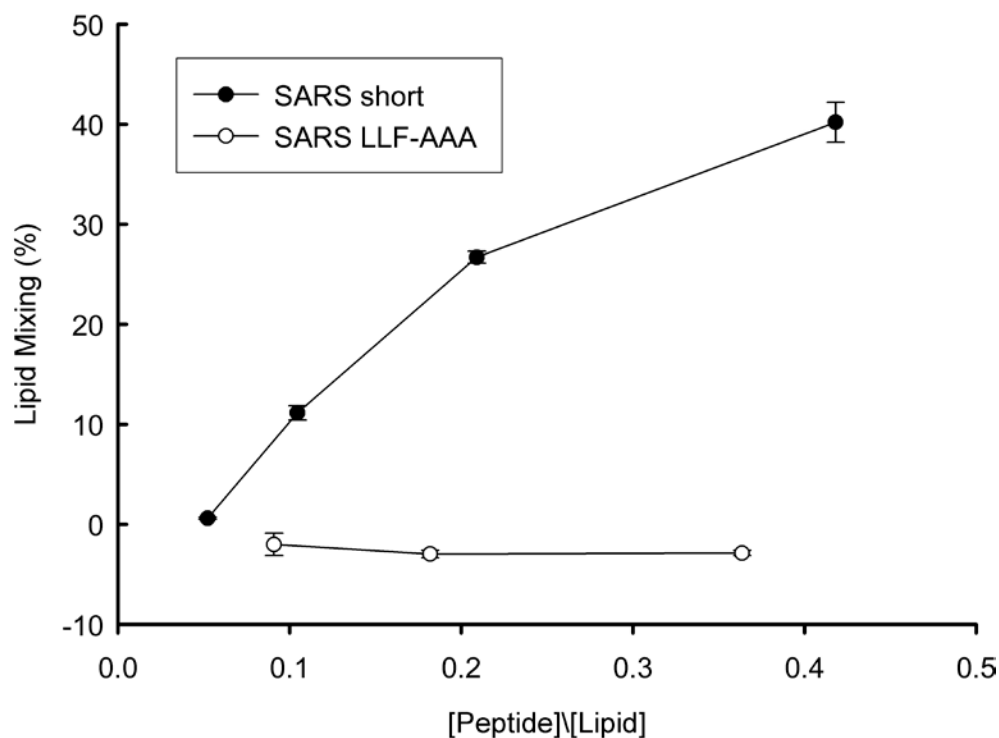


Figure 3.6 Lipid mixing by a short SARS-CoV S2 fusion peptide and a modified fusion peptide. The SARS short fusion peptide was able to mediate fusion at pH 7. Mutation of the core residues identified by cell-based assays, LLF, to AAA destroyed the ability of the peptide to mediate lipid mixing at pH 7.

### Feline Coronaviruses (FCoVs)

To advance our theory about the identity of the coronavirus fusion peptide, we also examined the hypothesized feline coronavirus fusion peptides. The FECV spike protein contains an analogous S2' cleavage site and fusion peptide as was discussed in Figure 3.3 and 3.4 and the exact sequence of the synthetic peptide is shown in Table 3.1. However, FIPV has a mutation in its spike protein at the S2' position which could shift the cleavage site as depicted in figure 3.3 and discussed previously. In order to address this issue we characterized two peptides to represent the potential fusion peptide of FIPV, one with a tyrosine and glycine residue at the N-terminus and

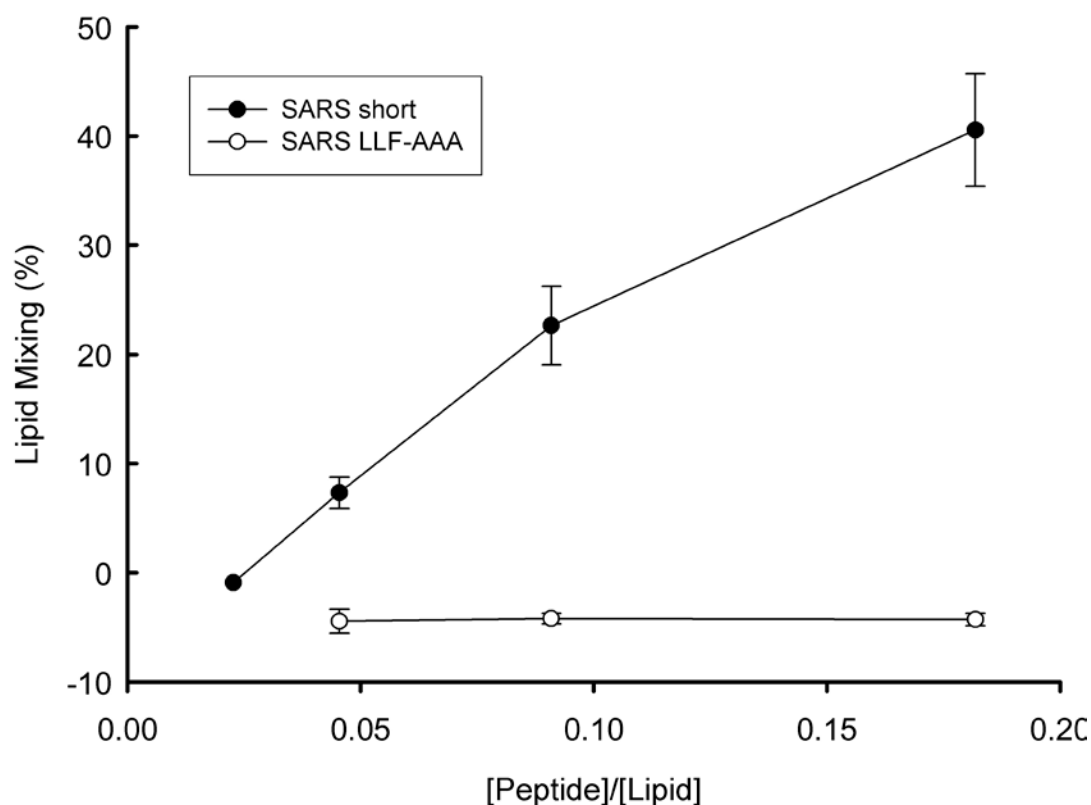


Figure 3.7 Lipid mixing by a short SARS-CoV S2 fusion peptide and a modified fusion peptide. The SARS short fusion peptide was able to mediate fusion at pH 5. Mutation of the core residues identified by cell-based assays, LLF, to AAA destroyed the ability of the peptide to mediate lipid mixing at pH 5.

a separate peptide with a glycine residue at the N-terminus. We carried out lipid mixing experiments on all of these possible feline coronavirus fusion peptides to better understand the effect of the N-terminal residues. The lipid mixing experiments were performed under similar conditions as the IBV and SARS experiments except that the composition of the liposomes was 4:1 POPG: cholesterol instead of 1:3:1 POPC:POPS:cholesterol. This change in liposome composition was necessary because the FECV peptide did not mediate lipid mixing with the composition used in the SARS and IBV experiments. Figure 3.8 shows the extent of lipid mixing, as

monitored by an increase in NBD fluorescence in the FRET-based lipid mixing assay, of the three feline coronavirus fusion peptides at pH 5. The fusion peptide of FIPV was able to mediate a higher extent of lipid mixing at a lower concentration of peptide than the FIPV-Y and FECV peptides. The ability of the FIPV-Y and FECV fusion peptides to promote lipid mixing at pH 5 is almost identical.

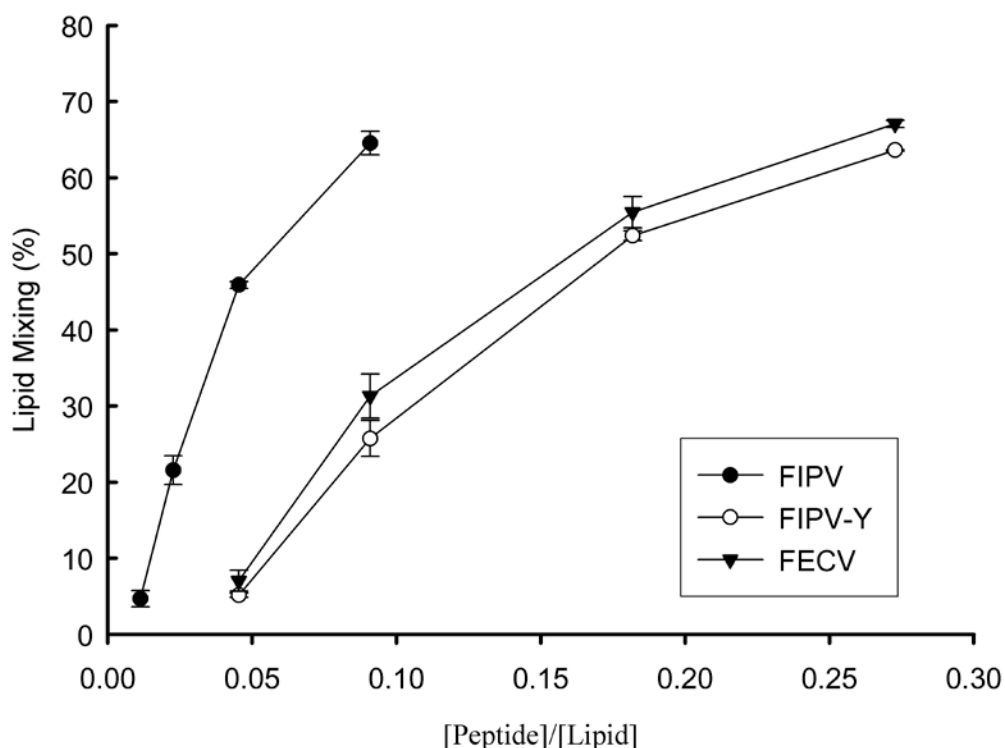


Figure 3.8 The FCoV S2 fusion peptides induce lipid mixing in liposomes at pH 5.

Extent of lipid mixing was determined by varying the ratios of FCoV S2 fusion peptides with total concentration of labeled and unlabeled 4:1 POPG:cholesterol liposomes at pH 5. The kinetics of mixing was followed by monitoring fluorescence intensity at 530nm upon addition of peptide. Each data point is averaged from three individual assays and error bars represent standard deviation of the mean.

We also examined the behavior of the feline coronaviruses at pH 7 as shown in



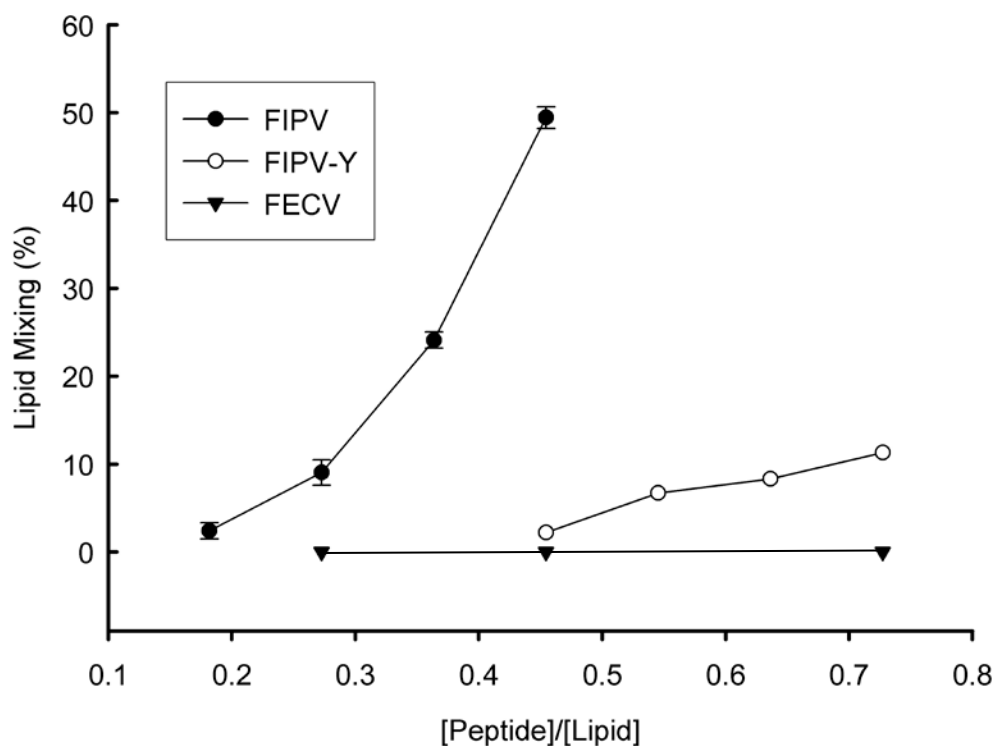


Figure 3.9. The FIPV fusion peptide is still able to promote lipid mixing, although at a higher concentration of peptide than was required at pH 5. The FIPV-Y peptide is also able to minimally promote lipid mixing at extremely high concentrations of peptide.

Figure 3.9 The FCoV S2 fusion peptides induce lipid mixing in liposomes at pH 7.

Extent of lipid mixing was determined by varying the ratios of FCoV S2 fusion peptides with total concentration of labeled and unlabeled 4:1 POPG:cholesterol liposomes at pH 7. The kinetics of mixing was followed by monitoring fluorescence intensity at 530nm upon addition of peptide.

Finally, the FECV fusion peptide is unable to promote lipid mixing even at the highest concentrations of peptide.

## ***Discussion***

The fusion protein of a virus and its fusion active component, the fusion peptide, are of central importance in the life cycle of a virus. Although much has been learned about fusion proteins and peptides for several viruses, including influenza, much about them is still a mystery. Here, we have shown evidence that a peptide in the S2 region in coronavirus fusion proteins, which is C-terminal to a cleavage site, has the potential to be the coronavirus fusion peptide.

Our experiments with the IBV fusion peptide in the FRET-based lipid mixing assay showed that the ability of the peptide to promote lipid mixing was dependent on pH. IBV is known to be a pH dependent virus [35], and probably enters cells through the endosomal pathway. The fact that the fusion peptide is active at low pH and inactive at neutral pH could contribute to the pH dependence of the virus. It is possible that the IBV fusion peptide adopts the correct structure to promote lipid mixing at low pH and cannot obtain that structure at neutral pH. When comparing the IBV peptide to the SARS peptides it is not obvious why the behavior of these two peptides is so different at pH 7. Although it is reasonable to think that the overall charge of the IBV, SARS and SARS short peptides could explain the pH dependence, closer examination of that hypothesis reveals deeper complexity. The IBV synthetic peptide contains 3 charged residues (excluding the linker region) and is closest in length to SARS short. SARS short contains 2 charged residues (excluding linker) and displays a similar ability to promote lipid mixing at pH 5 and 7. From this comparison one might assume that a decrease in charged residues leads to a decrease in pH dependence of lipid mixing. However, the longer synthetic version of the SARS fusion peptide has 5 charged residues (excluding linker), two more than the IBV synthetic peptide, and is able to promote fusion at pH 7, disproving the previous hypothesis. To completely rule out the effect of charge, an IBV and SARS synthetic

peptide of the same length would have to be subjected to the lipid mixing assay. Another possible explanation could be the influence of the charged residues on the fusion peptide secondary structure. The exact position of the charged residues within the IBV and SARS fusion peptides are different. It is possible that those charged residues are affecting the structure of those peptides differently as a function of position and pH. To further explore this hypothesis, structural studies of several modified fusion peptides which compare mutations at these charged residues at pH 5 and 7 would be required. Overall, the fact that the ability of the IBV fusion peptide to promote lipid mixing is pH dependent confirms that it acts as expected in different environments, lending strength to the argument that this region is the fusion peptide for IBV.

Experiments characterizing the ability of our proposed SARS fusion peptide to promote lipid mixing showed that it is able to do so at pH 5 and 7. This result is reasonable considering that SARS can infect cells by fusion at the cell membrane or in the endocytic pathway [36]. Additionally, comparison of lipid mixing experiments with the short SARS peptide and the long SARS peptide, show that the short peptide is less sensitive to pH than the long version of the SARS fusion peptide. From these observations it is possible to hypothesize that the residues SFIEDLLF are responsible for the lipid mixing activity. The residues NKVTLADAGFMKQY, which are present in the long peptide but not the short version, seem to have a pH dependent influence on lipid mixing which can be explained by the presence of charged residues in that portion of the peptide. Further experiments showed that mutation of the core residues of the synthetic SARS fusion peptide, LLF, to alanine destroyed the peptide's ability to mediate lipid mixing. This result is similar to that seen in the cell based assays, and highlights the close agreement between the in vitro assay and the function of the fusion peptide in the context of the virus.

Experiments with the proposed feline coronavirus fusion peptides showed that the position of the S2' cleavage site and therefore the sequence of the fusion peptide had a drastic effect on the peptide's ability to mediate lipid mixing. The fusion peptide for FIPV was able to promote lipid mixing at both pH 5 and 7. The proposed fusion peptide for FECV was able to promote lipid mixing at pH 5 but completely unable to aid in lipid mixing at pH 7. The intermediate peptide, called FIPV-Y, which had an N-terminal glycine residue was able to promote lipid mixing at pH 5 to a similar extent as the FECV fusion peptide but was only able to promote lipid mixing at pH 7 when a large concentration of fusion peptide was used. These results indicate that the N-terminal tyrosine and glycine of the FIPV fusion peptide have a pH dependent effect on lipid mixing. At pH 5, the tyrosine residue increases the extent of lipid mixing, but at pH 7 the tyrosine and glycine residues cooperatively increase the ability of the fusion peptide to promote lipid mixing. Without these crucial residues, the fusion peptide is unable to promote lipid mixing at pH 7. It is known that FECV is pH dependent, while FIPV is not [37], and the behavior of the fusion peptide could contribute to the difference in pH dependence and even the difference in pathogenicity of the two viruses. More explicitly, if the mutation in S2' cleavage site leads to the addition of a tyrosine and glycine residue to the N-terminal end of the fusion peptide, FIPV would be able to fuse with cells at neutral pH at the cell surface. This could allow FIPV to expand its cell tropism to monocytes because it can now avoid the harsh environment of the monocyte endosomal pathway. Therefore, FIPV can infect a larger range of cell types and, as a consequence, is a more widely spread and lethal virus.

### ***Conclusions***

The mechanism of fusion is relevant in many critical processes effecting human life from the release of neurotransmitters to the entry of viruses into cells. In

our efforts to understand this process, viral fusion proteins have been identified as a key component of viral entry and a model to help us understand membrane fusion. Despite the fact that much has been learned, there is still more to be discovered. It is hoped that by identifying the fusion peptide, the part of the viral fusion protein that is inserted into the host membrane, we might be able to stop or restrict this process, leading to a new therapeutic strategy.

A significant amount of experimental evidence has been revealed suggesting that the region C-terminal to the S2' cleavage site is the fusion peptide in many if not all of the members of the coronavirus family. Here, we provided evidence showing that this region is able to associate with lipid membranes and mediate lipid mixing as would be expected of a fusion peptide. Additionally, many of these peptides did so in a way that parallels the pH dependence of the virus in question. This hypothesized fusion peptide possesses several characteristics that make it a better candidate than the previously suggested regions of the S protein. Most important is the fact that our hypothesized fusion peptide is highly conserved among the coronaviruses, unlike WWI and WWII. Also, our fusion peptide is N-terminal to an established protease cleavage site, and the previously suggested fusion peptides are not. Granting that these experiments do not prove this region is the fusion peptides, more work must be done. Ideally, an experiment proving that this region is actually inserted into the host membrane during infection and that it causes the formation of the fusion pore is needed. It is possible that this could be accomplished through the use of a fusion peptide antibody. If an antibody to the coronavirus fusion peptide was shown to have a deleterious effect on infection, it could be used to show that the fusion peptide is exposed after an environmental trigger and that it is inserted into the host cell membrane. Perhaps antibody experiments as described in combination with all of the other evidence gathered thus far would be enough to prove the region N-terminal to

the S2' cleavage site is the fusion peptide. Alternatively, membrane cross-linking experiments could be used to show that the region in question is inserted into the membrane after being triggered by environmental conditions. A membrane cross-linking experiment would also be very persuasive in building a case for the identity of the coronavirus fusion peptide. Due to the controversy surrounding the identification of the coronavirus fusion peptides, as well as the need to understand these potentially devastating human pathogens, it is certain that experiments that conclusively identify the fusion peptide are actively being sought out.

## REFERENCES

1. Durell, S.R., et al., *What studies of fusion peptides tell us about viral envelope glycoprotein-mediated membrane fusion*. Molecular Membrane Biology, 1997. **14**(3): p. 97-112.
2. White, J.M., et al., *Structures and mechanisms of viral membrane fusion proteins: Multiple variations on a common theme*. Critical Reviews in Biochemistry and Molecular Biology, 2008. **43**(3): p. 189-219.
3. Skehel, J.J. and D.C. Wiley, *Receptor binding and membrane fusion in virus entry: The influenza hemagglutinin*. Annual Review of Biochemistry, 2000. **69**: p. 531-569.
4. Tamm, L.K., et al., *Structure and function of membrane fusion peptides*. Biopolymers, 2002. **66**(4): p. 249-260.
5. Wilson, I.A., J.J. Skehel, and D.C. Wiley, *Structure of the Hemagglutinin Membrane Glycoprotein of Influenza-virus at 3-A Resolution*. Nature, 1981. **289**(5796): p. 366-373.
6. Roche, S., *Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G (vol 313, pg 187, 2006)*. Science, 2006. **313**(5792): p. 1389-1389.
7. Roche, S., et al., *Structure of the prefusion form of the vesicular stomatitis virus glycoprotein G*. Science, 2007. **315**(5813): p. 843-848.
8. Peuvot, J., et al., *Are the fusion processes involved in birth, life and death of the cell depending on tilted insertion of peptides into membranes?* Journal of Theoretical Biology, 1999. **198**(2): p. 173-181.
9. Durrer, P., et al., *Photolabeling Identifies a Putative Fusion Domain in the Envelope Glycoprotein of Rabies and Vesicular Stomatitis Viruses*. Journal of

- Biological Chemistry, 1995. **270**(29): p. 17575-17581.
10. Durrer, P., et al., *H<sup>+</sup>-induced membrane insertion of influenza virus hemagglutinin involves the HA2 amino-terminal fusion peptide but not the coiled coil region*. Journal of Biological Chemistry, 1996. **271**(23): p. 13417-13421.
  11. Roche, S. and Y. Gaudin, *Pathway of virus-induced membrane fusion studied with liposomes*, in *Liposomes, Pt B*. 2003. p. 392-407.
  12. Nieva, J.L. and A. Agirre, *Are fusion peptides a good model to study viral cell fusion?* Biochimica Et Biophysica Acta-Biomembranes, 2003. **1614**(1): p. 104-115.
  13. Duzgunes, N., *Fluorescence assays for liposome fusion*, in *Liposomes, Pt B*. 2003. p. 260-274.
  14. Struck, D.K., D. Hoekstra, and R.E. Pagano, *Use of resonance Energy transfer to Monitor Membrane Fusion*. Biochemistry, 1981. **20**(14): p. 4093-4099.
  15. Maier, O., V. Oberle, and D. Hoekstra, *Fluorescent lipid probes: some properties and applications (a review)*. Chemistry and Physics of Lipids, 2002. **116**(1-2): p. 3-18.
  16. Blumenthal, R., et al., *Fluorescent lipid probes in the study of viral membrane fusion*. Chemistry and Physics of Lipids, 2002. **116**(1-2): p. 39-55.
  17. Peisajovich, S.G. and Y. Shai, *Liposomes in identification and characterization of viral fusogenic peptides*, in *Liposomes, Pt B*. 2003. p. 361-373.
  18. Han, X. and L.K. Tamm, *A host-guest system to study structure-function relationships of membrane fusion peptides*. Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(24): p. 13097-13102.



19. Korte, T., et al., *Role of the Glu residues of the influenza hemagglutinin fusion peptide in the pH dependence of fusion activity*. Virology, 2001. **289**(2): p. 353-361.
20. Pritsker, M., et al., *Effect of nonpolar substitutions of the conserved Phe(11) in the fusion peptide of HIV-1 gp41 on its function, structure, and organization in membranes*. Biochemistry, 1999. **38**(35): p. 11359-11371.
21. Nieva, J.L., et al., *Interaction of the HIV-1 Fusion Peptide with Phospholipid-Vesicles - Different Structural Requirements for Fusion and Leakage*. Biochemistry, 1994. **33**(11): p. 3201-3209.
22. Ghosh, J.K., S.G. Peisajovich, and Y. Shai, *Sendai virus internal fusion peptide: Structural and functional characterization and a plausible mode of viral entry inhibition*. Biochemistry, 2000. **39**(38): p. 11581-11592.
23. Gomara, M.J., et al., *Roles of a conserved proline in the internal fusion peptide of Ebola glycoprotein*. Febs Letters, 2004. **569**(1-3): p. 261-266.
24. Ruiz-Arguello, M.B., et al., *Phosphatidylinositol-dependent membrane fusion induced by a putative fusogenic sequence of Ebola virus*. Journal of Virology, 1998. **72**(3): p. 1775-1781.
25. Sainz, B., et al., *Identification and characterization of the putative fusion peptide of the severe acute respiratory syndrome-associated coronavirus spike protein*. Journal of Virology, 2005. **79**(11): p. 7195-7206.
26. Guillen, J., et al., *Interaction of a peptide from the pre-transmembrane domain of the severe acute respiratory syndrome coronavirus spike protein with phospholipid membranes*. Journal of Physical Chemistry B, 2007. **111**(49): p. 13714-13725.
27. Lai, M.M.C.a.H., K. V., *Coronaviridae: The viruses and their replication*, in *Field's Virology*, D. Knipe, Editor. 2001, Lippincott, Williams, and Wilkins:

- Philadelphia. p. 1163-1185.
28. Hofmann, H. and S. Pohlmann, *Cellular entry of the SARS coronavirus*. Trends in Microbiology, 2004. **12**(10): p. 466-472.
  29. Yang, Z.Y., et al., *pH-dependent entry of Severe acute respiratory syndrome coronavirus is mediated by the spike glycoprotein and enhanced by dendritic cell transfer through DC-SIGN*. Journal of Virology, 2004. **78**(11): p. 5642-5650.
  30. Regan, A.D. and G.R. Whittaker, *Utilization of DC-SIGN for Entry of Feline Coronaviruses into Host Cells*. Journal of Virology, 2008. **82**(23): p. 11992-11996.
  31. Guillen, J., et al., *Structural and dynamic characterization of the interaction of the putative fusion peptide of the S2SARS-CoV virus protein with lipid membranes*. Journal of Physical Chemistry B, 2008. **112**(23): p. 6997-7007.
  32. Guillen, J., P.K.J. Kinnunen, and J. Villalain, *Membrane insertion of the three main membranotropic sequences from SARS-CoV S2 glycoprotein*. Biochimica Et Biophysica Acta-Biomembranes, 2008. **1778**(12): p. 2765-2774.
  33. Belouzard, S., Chu, V. C., and Whittaker, G. R., *Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites*. Proceedings of the National Academy of Sciences of the United States of America, 2009. **Early Edition**.
  34. Madu, I., Roth, S. L., Belouzard, S. and Whittaker, G. R., *Characterization of a novel fusion peptide within the SARS coronavirus spike protein S2 domain*. Journal of Virology, 2009. **Accepted**.
  35. Chu, V.C., et al., *The avian coronavirus infectious bronchitis virus undergoes direct low-pH-dependent fusion activation during entry into host cells*. Journal of Virology, 2006. **80**(7): p. 3180-3188.

36. Matsuyama, S., et al., *Protease-mediated enhancement of severe acute respiratory syndrome coronavirus infection*. Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(35): p. 12543-12547.
37. Regan, A.D., et al., *Differential role for low pH and cathepsin-mediated cleavage of the viral spike protein during entry of serotype II feline coronaviruses*. Veterinary Microbiology, 2008. **132**(3-4): p. 235-248.

## CHAPTER 4

### SUMMARY AND CONCLUSIONS

Viral entry is an extremely active and important area of research. By investigating the first interactions of viruses with their host cells not only do we learn possible ways to prevent infection, but we also learn about natural cell processes. The studies outlined in this thesis contribute to our understanding of the entry mechanism of VSV, the fusion mechanism of coronaviruses.

#### *Summary and Conclusions*

Vesicular stomatitis virus has been used as a model of virus-host cell interactions for an extended period of time. Although VSV is well studied, there are details of its entry mechanism which are still unknown. In chapter 2, we investigated the interaction of the glycoprotein of VSV with the endosome specific lipid LBPA. We found that the VSV G protein interacted with LBPA in a manner that accelerated fusion. Additionally, this interaction was specific to VSV G and the presence of LBPA did not accelerate the fusion rate of other viruses. These observations provide an explanation for a two-step fusion and infection mechanism where VSV fuses into the internal vesicles of MVBs followed by infection from the late endosome. Due to the fact that LBPA is present in the internal vesicles of the MVBs our evidence supports the hypothesis by showing that the VSV G protein interacts with LBPA which encourages fusion with these vesicles. The two step entry mechanism proposed for VSV and supported by our experiments is unique and, with further study, could provide insight into the regulation of the endosomal pathway.

The merging of the viral envelope with the host cell membrane is mediated by the viral fusion protein. Specifically, the fusion peptide is inserted into the host cell membrane and functions to destabilize the host cell membrane and bring the viral and host cell membrane close enough together to promote fusion. In chapter 3, we

characterize the ability to promote lipid mixing of our proposed coronavirus fusion peptides. We show that the proposed fusion peptides have the ability to promote lipid mixing and therefore could be inserted into the host cell membrane and cause the formation of the fusion pore. Additionally, we see evidence in our lipid mixing assay that some of the fusion peptides promote lipid mixing in a pH dependent manner which is analogous to the pH requirements of the virus. These correlations suggest that the ability of the fusion peptide to promote lipid mixing could contribute to the pH dependence of this family of viruses. Additionally, we observed that mutations to the important residues of our fusion peptides abolished its ability to promote lipid mixing. Indicating that small changes in our fusion peptide can have a devastating effect on its activity which correlates with cell based assays.

### ***Future Directions***

Identification of an interaction between the VSV G protein and the lipid LBPA is just the first step towards supporting the two step mechanism of VSV entry. Future experiments could include extensive mutagenesis of the VSV protein in an attempt to determine which residues are responsible for this effect. Additionally, it would be interesting to attempt to manipulate the conditions of cells and determine the effect of different stressors on the entry pathway taken by VSV. It might be possible to find an environmental factor that made one pathway or the other more favorable. These lines or research would help us further understand this novel entry pathway.

A significant amount of experimental evidence has been revealed suggesting that the region C-terminal to the S2' cleavage site is the fusion peptide of the coronavirus family. However, we have still not conclusively proven that this region is the fusion peptide. As stipulated previously, an experiment must be carried out showing that cleavage at S2' exposes our proposed fusion peptide. One possible way to do that would be to make an antibody to the coronavirus fusion peptide. Using this

antibody it would be possible to establish whether the fusion peptide was exposed upon cleavage. Additionally, if it could be shown that the antibody blocks infection, this evidence could indicate that the antibody is preventing the insertion of the fusion peptide into the host cell membrane. In addition, it is necessary to show that the proposed fusion peptide promotes the formation of the fusion pore. In order to prove this, a membrane cross linking experiment could be employed to show that our proposed fusion peptide is specifically being inserted into the host cell membrane. As a direct result of the high level of interest in coronaviruses and their potential as significant human pathogens as proven by the SARS outbreak of 2003, I believe that all of these methods will be utilized to further our understanding of the coronavirus fusion machinery.

## APPENDIX 1

### THE EFFECT OF N-GLYCOSYLATION OF HUMAN CELLS ON INFLUENZA INFECTION

#### *Introduction*

#### **Influenza**

Influenza is an enveloped virus with a negative single-stranded RNA genome. There are 8 segments to the genome which encode 10 or 11 proteins. These proteins include PB1, PB2 and PA which form the transcriptase complex. The three transmembrane proteins HA (hemagglutinin), NA (neuraminidase) and M2 (an ion channel) are inserted through the lipid bilayer and are involved in entry and budding. The M1 or matrix protein underlies the lipid bilayer. The nucleocapsid protein (NP) associates with the viral RNA. Finally, there are two non-structural proteins NS1 and NS2. NS1 is found in infected cells but is not a major part of the virion itself, and NS2 is found associated with virions in small quantities. The hemagglutinin protein plays a key role in the binding and entry of the influenza virus. The HA protein is cleaved into two parts during the viral life cycle which are termed HA1 and HA2. HA 1 is responsible for receptor binding and HA2 contains the fusion machinery. A crystal structure is available which has greatly added to our understanding of viral fusion proteins.

The first stage in a viral lifecycle is the binding of the virus to the membrane of the host cell. The influenza HA protein, more specifically HA1, is known to bind to sialic acid which triggers receptor mediated endocytosis in both a clathrin dependent and independent manner. A crystal structure showing the binding hemagglutinin with its receptor sialic acid has been solved [1]. One can see that the sialic acid binds in a shallow depression at the top of the HA molecule. The residues that the receptor

interacts with are conserved and not prone to antigenic drift. NMR binding studies show that the binding of a single sialic acid residue is a low affinity event. It has been concluded that several interactions are needed for strong virus binding. The conformation of the sialic acid linkage controls the species tropism of the influenza strain. Sialic acid can either be bound to the hydroxyl group of the third carbon of a galactose molecule to form an  $\alpha$ 2-3 linkage or to the hydroxyl group of the 6 carbon to form an  $\alpha$ 2-6 linkage. Influenza strains that infect avian species have a preference for  $\alpha$ 2-3 linked sialic acid moieties. Strains of influenza that infect humans have a preference for  $\alpha$ 2-6 linkage. It has been observed that the mucins that are present in the upper respiratory tract of humans protect them from avian influenza strains by containing many sialic acids in the  $\alpha$ 2-3 linkage conformation which compete with cell surface receptors for viral binding. Inhibition of the interaction between influenza HA and the cell surface sialic acid residues is being investigated as a possible therapeutic strategy [2]. As mentioned previously, the binding between one sialic acid residues and HA is weak, so a multivalent sialyloligosaccharide is required for a significant effect. Although no pharmaceutical is yet available, multivalent sialyloligosaccharides and multivalent sialylmimetics show promise as new therapies.

Sialic acid residues can be found on both glycolipids and glycoproteins. It has been established that gangliosides, or sialic acid residue containing glycolipids, are not essential for influenza virus infection [3]. In this study, cells lacking gangliosides called GM-95 cells were infected with influenza virus. Although the sensitivity of the cells to several influenza strains were 2-4 times lower than that of wild types cells, it was clear that the lack of gangliosides did not prevent productive infection. To determine whether a glycoprotein could be the receptor for influenza, a series of experiments were undertaken with cells that had a deficiency in receptor sialo-N-glycans (Lec 1 cells) [4]. These studies show that although Lec 1 cells are still able to



undergo virus binding, fusion, and replication. However, the virus is not able to initiate host cell infection which indicates a block during entry after the initial virus binding. The authors further hypothesize that an N-linked glycoprotein is the receptor for influenza virus.

### **MGAT1**

UDP-GlcNAc:α3-D-mannoside β1,2-N-acetylglucosaminyltransferase I (encoded by the gene MGAT1) controls the synthesis of hybrid, complex, and paucimannose N-glycans in all eukaryotes [5]. It initiates the synthesis of these complex glycans by catalyzing the transfer of GlcNAc from UDPGlcNAc to the oligomannosyl acceptor Man<sub>5</sub>GlcNAc<sub>2</sub>-Asn [6]. The oligomannosyl acceptor is found on approximately 70% of cell-surface receptors and transporters with the motif Asn-X/Ser/Thr (X≠ Pro) that are exposed to the lumen of the endoplasmic reticulum [7]. The role of these complex N-glycans is so vital during development that an absence of the MGAT1 gene in mouse embryos is lethal [8], although it does not seem to be vital to life as evidenced by the existence of cells lines without a functioning MGAT1 gene [9].

### **Short Interfering RNA (siRNA)**

Short interfering RNA are small (approximately 20-30 nucleotide base) non-coding RNA sequences that can regulate genes and genomes [10]. A phenomenon related to siRNA, micro RNA (miRNA), was discovered in 1993 as an endogenous mechanism to control gene expression during development [11]. A key study in 2004 found that exogenous dsRNA could be used to invoke similar mechanisms and specifically silence genes [12, 13]. The realization that dsRNA could be introduced to a cell exogenously to effect the expression of genes has had a profound impact on research in the biological sciences. siRNAs are not only found during the developmental stages as they were originally found, they can also be derived from

viruses, centromeres, transposons, and endogenous genomic loci. During the endogenous process, siRNAs are cut from the ends of longer double stranded RNAs by a protein called Dicer. These siRNAs then induced the formation of a silencing complex named RISC. In humans a RISC is composed of Dicer, TRBP and Ago-2 [10]. This complex binds the siRNA and loads it into Ago-2. Ago-2 cleaves the passenger strand, or the strand of the RNA duplex that will not be used to base pair to the target, and it is discarded. After these steps have been accomplished the complex is active and can begin to influence gene expression. Silencing is achieved by the active RISC by finding the target gene through Watson-Crick base pairing. When the target gene has been located, it is degraded by Ago [10]. Another mechanism that RISC uses to modulate transcription is to associate with newly transcribed RNA, again through Watson-Crick base pairing. This association promotes histone methylation and the formation of heterochromatin, a compact DNA structure. The formation of heterochromatin effectively down regulates the transcription of the target gene [10].

To determine if the presence of N-linked glycoproteins were also important during the infection of human cells with influenza, I attempted to knock down the human equivalent of GnT1, called MGAT1. GnT1 is the enzyme in hamster cells that was disrupted in the Lec 1 cells rendering them incapable of synthesizing mature N-linked glycoproteins. In order to achieve this goal, siRNA was utilized for transient modulation of gene expression and effect was monitored by Western blot, lectin binding and influenza infection. An attempt to establish a cell line stably expressing siRNA to knock down the MGAT1 gene was also made.

## ***Materials and Methods***

### **Cells and virus**

Cells utilized during this study, A549, HeLa, MDBK, and HEK-293T were obtained from ATCC. The A549 cells were grown in Ham's F12 media with 10%

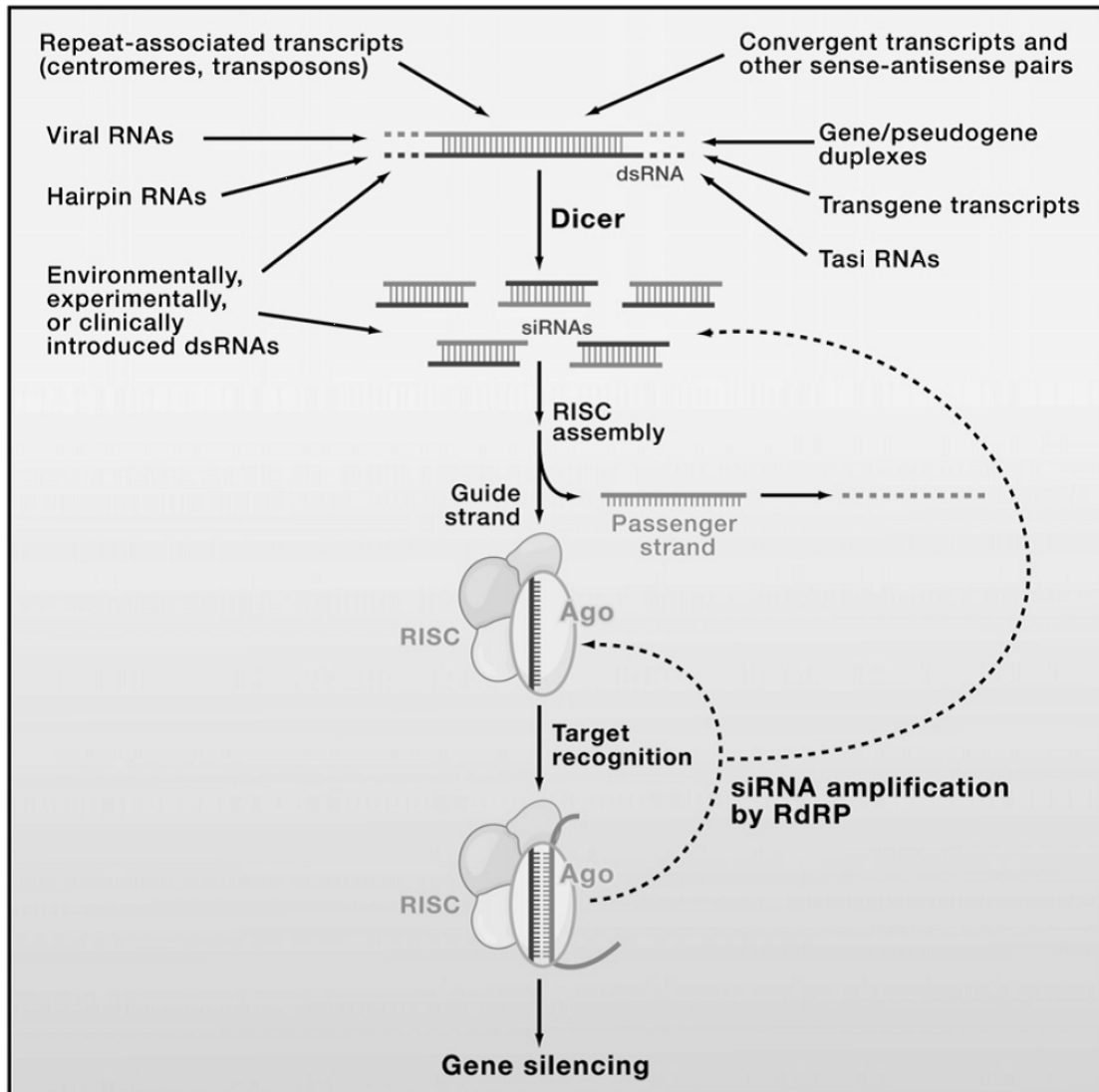


Figure A1.1 Overview of the siRNA pathway (Carthew & Sontheimer, *Cell*, 2009).

fetal bovine serum, 100 U/mL penicillin and 100 U/mL streptomycin. All other cell types were grown in DMEM with 10 % fetal bovine serum, 10 mM HEPES, 100 U/mL penicillin and 100 U/mL streptomycin. Influenza (A/WSN/33) was obtained from ATCC and grown in MDBK cells inoculated at 0.01 MOI and harvested in DMEM containing 2% FBS, 10 mM HEPES, 100 U/mL penicillin and 100 U/mL streptomycin. All cell culture reagents were purchased from Mediatech Inc.

#### **MGAT1 siRNA knockdown and influenza infection**

A549 cells were trypsinized and plated in 24-well plates in Ham's F12 with 10% FBS, 100 U/mL penicillin and 100 U/mL streptomycin, 24 hours prior to transfection. A 25 pmol amount of the siRNA indicated was transfected with 0.5  $\mu$ L of Lipofectamine 2000 (Invitrogen). The RNA oligomers used were obtained from Ambion, and the Silencer Pre-designed siRNA database. The sequences were obtained in annealed form and denoted as: 11388 (sense) GGCCUAUGACCGAGAUUUCtt (antisense) GAAAUCUCGGUCAUAGGCCtc; 11481 (sense) GGGACAGCUUCAAGGCUUUt (antisense) AAAGCCUUGAAGCUGUCCctg; 11566 (sense) GGCUGAAUUGUCUGAAAAAtt (antisense) UUUUUCAGACAAUUCAGCCtt; and 289360 (sense) GAUCUCAAGAACGAUGACctt (antisense) GGUCAUCGUUCUUGAGAUCtc. Approximately 48 hours after transfection, cells were assayed by Western blot for expression of MGAT 1 using a rabbit anti-MGAT 1 polyclonal antibody (Abgent) or infected with influenza A/WSN/33.

#### **MGAT1 stability test**

A549 cells were placed in a 24 well plate at 50% confluence. After reaching 70-80% confluence cycloheximide was added to the media at a concentration of 5  $\mu$ g/mL. At regular intervals, cells were lysed and the lysate prepared for Western blot analysis.

#### **Plasmid construction and stable cell line generation**

Generation of stable cell lines utilized the pSuper.retro.neo system (Oligoengine). Briefly, DNA oligomers were obtained (Oligoengine) based on the siRNA sequences 11481 and 289360 and were annealed. The vector pSuper.retro.neo was linearized using BglII and HindIII and the DNA duplexes were ligated to the vector. After bacterial transformation, clones were isolated and sequenced to verify the nucleotide sequence. The pSuper.retro plasmids utilizing the 11481 sequence are

denoted as 24 and 26 and the plasmids utilizing the 289360 sequences are denoted as 1

### **Results**

In the initial experiments, a group of siRNA targeted to the MGAT1 gene were transfected to achieve a transient knockdown of the gene. As shown in Figure A1.2, the level of MGAT1 expression after a single transformation of siRNA in A549 cells as determined by quantitative Western Blot was practically unchanged.

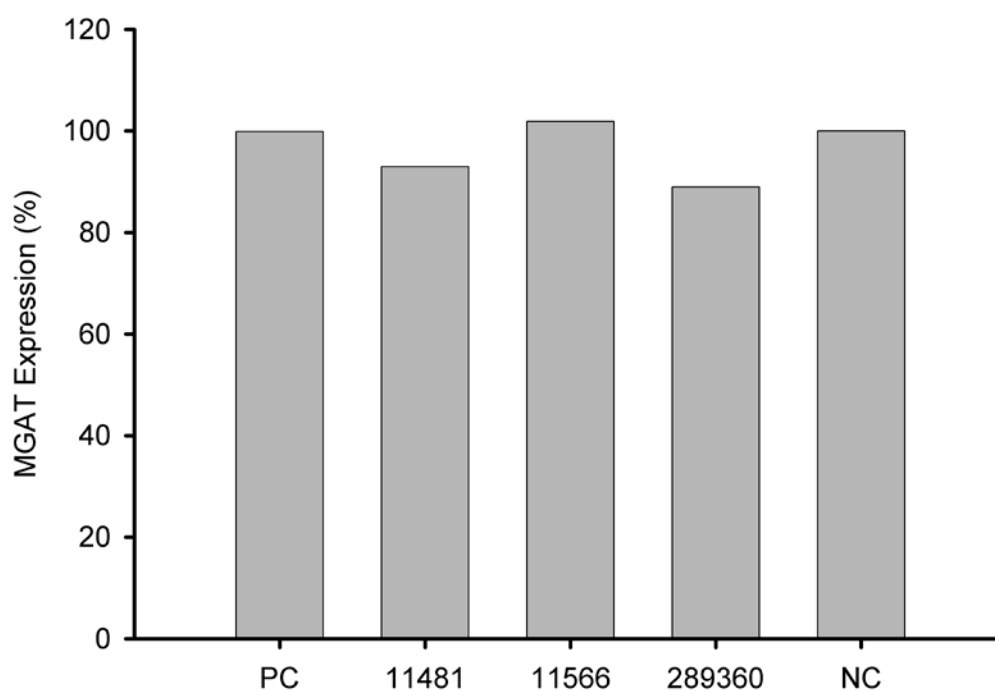


Figure A1.2 Expression of MGAT1 after a single transfection of siRNA in A549 cells. The level of MGAT1 expression was determined by quantitative Western blot.

Due to the fact that some proteins have longer half lives than others, it seemed from this experiment that the half life of MGAT1 was longer than the experimental time line.

In order to rectify that, a second transfection of siRNA was utilized as shown in Figure A1.3

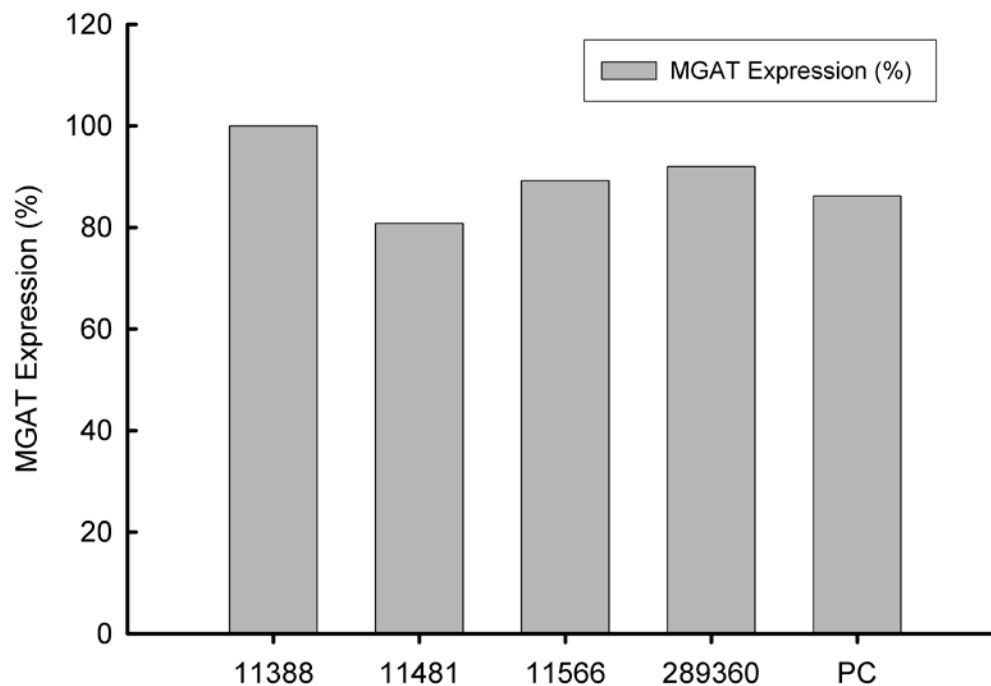


Figure A1.3 Expression of MGAT1 after two transfections of siRNA in A549 cells. The level of MGAT1 expression was determined by quantitative Western blot.

However, a second transfection of siRNA did not seem to have an appreciable effect on the expression level of MGAT1 by any of the siRNA duplexes tested. In order to determine whether a minimal amount of expression suppression on MGAT1 had an influence on the ability of influenza to infect A549 cells, an experiment was carried out that attempted to infect A549 cells that were transfected twice with siRNA, as shown in Figure A1.4. In an attempt to determine the half life of MGAT1, which could be unusually long due to its nature as a transmembrane protein present in the medial-Golgi, a stability time course with cycloheximide was attempted. Cycloheximide is a inhibitor of protein biosynthesis in eukaryotic organisms.

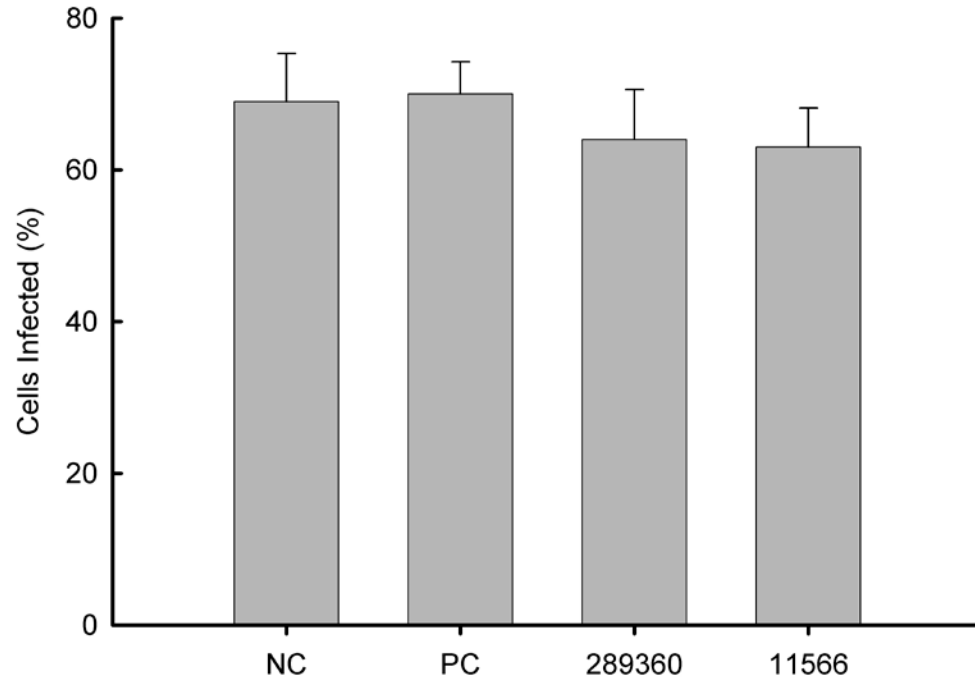


Figure A1.4 Percentage of A549 cells infected by influenza WSN/33 after two transfections of siRNA. NC and PC correspond to transfection of the negative control and positive control siRNA and are expected to have no effect on influenza infection.

It works by interfering with translocation during translation which blocks nascent peptide elongation. By adding cycloheximide to the media in cell culture, protein synthesis is inhibited and therefore, the protein expression level monitored over time by Western blot is due to protein present before the addition of cycloheximide. The results of a time course of the expression of MGAT1 in the presence of cycloheximide is shown in Figure A1.5.

### ***Discussion***

As can be seen in Figures A2.1, and A2. 2 transient transfection of our siRNA duplexes was incapable of significantly decreasing MGAT1 expression. At its greatest extent, the amount of knockdown was only 20%, which is not enough to have

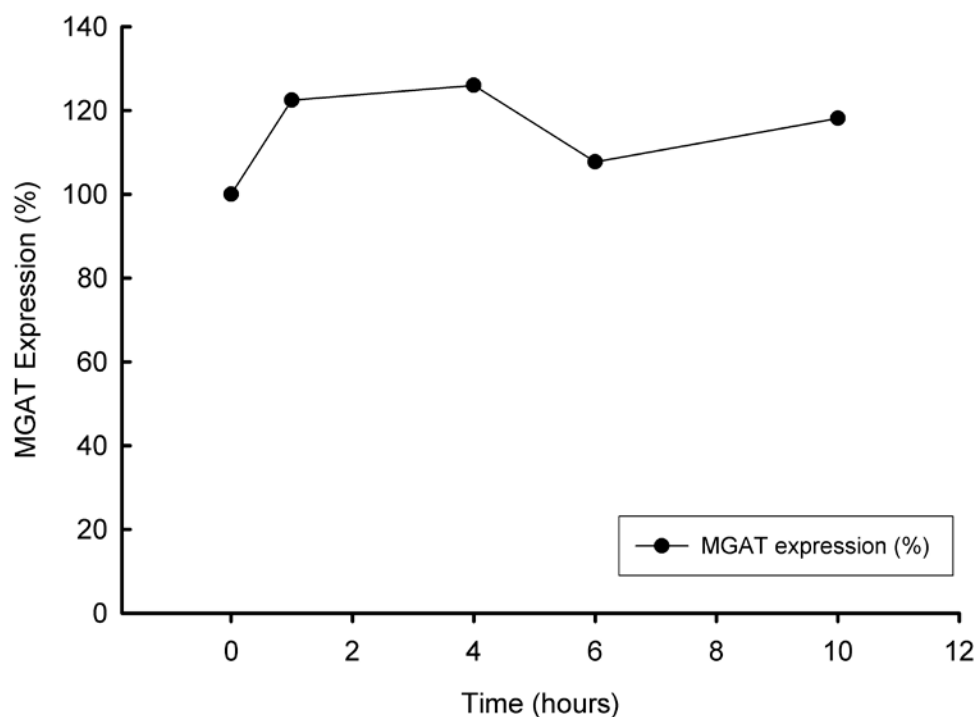


Figure A1.5 Effect of cycloheximide on MGAT1 expression over a period of 12 hours.

a biological effect. The lack of biological effect was confirmed by a lack of effect on fluorescein labeled lectin binding (data not shown.) Additionally, the transfection of siRNA had no effect on the infection of cells by influenza A/WSN/33. To determine the stability of MGAT1 A549 cells were exposed to cycloheximide and the amount of MGAT1 expressed over time was monitored. As shown in figure A2. 4 over a period of 12 hours no change in the amount of MGAT1 present in A549 cells was observed, indicating that the protein was exceptionally stable and might require the generation of a stable cell line to significantly be decreased with siRNA. The MGAT1 expression in four stably transfected cells lines are shown in figure A2.5. As in the transient transfection experiments, MGAT1 expression was only decreased by 20% in the most highly affected cell line. Fluorescein labeled lectin binding to these cells showed no



difference between their ability to bind lectins and wild type cells lectin binding ability.

### ***Conclusions***

Unfortunately, these experiments offer little insight to the effect of MGAT1 expression on influenza infection in human cells. There are several possible explanations for the difficulties encountered in these experiments. First, it is possible that the siRNA sequences used did not efficiently knock down the desired gene MGAT1. A recent literature search has revealed no reported accounts of successful knockdown of MGAT1 with these or any other siRNA sequences although successful studies have been done with *Caenorhabditis elegans* and perhaps the use of siRNA analogous to the sequences used in that study would be successful [5]. Additionally, these siRNA sequences could be toxic to the host cells, which would affect both transient and stable transfection. Secondly, the MGAT1 protein seems to be unusually stable, which contributes to the difficulty in decreasing its expression. I think future experiments should attempt to use different siRNA sequences.

## REFERENCES

1. Weis, W., et al., *Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid*. Nature, 1988. **333**(6172): p. 426-431.
2. Sun, X.L., *Recent anti-influenza strategies in multivalent sialyloligosaccharides and sialylmimetics approaches*. Current Medicinal Chemistry, 2007. **14**(21): p. 2304-2313.
3. Matrosovich, M., et al., *Gangliosides are not essential for influenza virus infection*. Glycoconjugate Journal, 2006. **23**(1-2): p. 107-113.
4. Chu, V.C. and G.R. Whittaker, *Influenza virus entry and infection require host cell N-linked glycoprotein*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(52): p. 18153-18158.
5. Shi, H., J. Tan, and H. Schachter, *N-glycans are involved in the response of Caenorhabditis elegans to bacterial pathogens*, in *Functional Glycomics*. 2006, Elsevier Academic Press Inc: San Diego. p. 359-389.
6. Kumar, R., et al., *Cloning and expression of N-acetylglucosaminyltransferase-I, the medial golgi transferase that initiates complex N-linked carbohydrate formation*. Proceedings of the National Academy of Sciences of the United States of America, 1990. **87**(24): p. 9948-9952.
7. Lau, K.S., et al., *Complex N-glycan number and degree of branching cooperate to regulate cell proliferation and differentiation*. Cell, 2007. **129**(1): p. 123-134.
8. Ye, Z.Y. and J.D. Marth, *N-glycan branching requirement in neuronal and postnatal viability*. Glycobiology, 2004. **14**(6): p. 547-558.
9. Kumar, R. and P. Stanley, *Transfection of a human-gene that corrects the Lec1 glycosylation defect - evidence for transfer of the structural gene for N-*

*acetylglucosaminyltransferase-I*. Molecular and Cellular Biology, 1989. **9**(12): p. 5713-5717.

10. Carthew, R.W. and E.J. Sontheimer, *Origins and Mechanisms of miRNAs and siRNAs*. Cell, 2009. **136**(4): p. 642-655.
11. Bartel, D.P., *MicroRNAs: Genomics, biogenesis, mechanism, and function*. Cell, 2004. **116**(2): p. 281-297.
12. Hutvagner, G., et al., *Sequence-specific inhibition of small RNA function*. Plos Biology, 2004. **2**(4): p. 465-475.
13. Mello, C.C. and D. Conte, *Revealing the world of RNA interference*. Nature, 2004. **431**(7006): p. 338-342.

## APPENDIX 2

### DIRECTED EVOLUTION OF STOFFEL FRAGMENT DNA POLYMERASE

This work was done in the laboratory of Dr. D. Tyler McQuade in the Department of Chemistry and Chemical Biology at Cornell University. It is presented here with his consent.

#### *Introduction*

##### **Directed Evolution**

Directed evolution refers to a class of techniques that can be used to modify enzyme properties. Mutant enzymes are used widely as reagents for chemical transformation, reagents for molecular biology, and as protein-based therapeutics [1]. According to a recent report, 60% of large scale commercial asymmetric syntheses involve at least one enzymatic or whole cell biotransformation [2]. Enzymatic reagents are used to make optically pure intermediates and products for the basic chemical, fine chemical, food science, agrochemical, and pharmaceutical industries [1]. Because wild-type enzymes often do not display the activity or the broad substrate scope needed for an academic or industrial application, these enzymes must first be mutated. Two methods are available to mutate enzymes: rational design and directed evolution. Rational design relies on close inspection of protein models/crystal structures or a detailed understanding of active site structure and function to engineer a protein's function [1]. This method works well in some cases, but can fail to alter behavior for many enzymes. In addition, rational design may not effectively refine an enzyme's activity, selectivity, etc. Directed evolution (DE) selects proteins with the desired trait from large libraries of mutants in successive rounds. DE improves activity, selectivity, and stability very effectively. DE also enables the discovery of important residues away from the active site as well as quick access to structure activity relationships.

DE is a very effective method for refining a trait already exhibited by an enzyme [3, 4]. DE relies on the creation of a library of proteins consisting of mutants with multiple randomly placed mutations,[5] random mutations of a selected domain,[6] or saturation mutagenesis of a select number of amino acids [7]. DE's power is revealed by high-throughput screening/selection [8]. Since many mutations are produced and selected in one round, DE's success is predicated on how the genotype and phenotype are linked (Figure A2.1) [9].

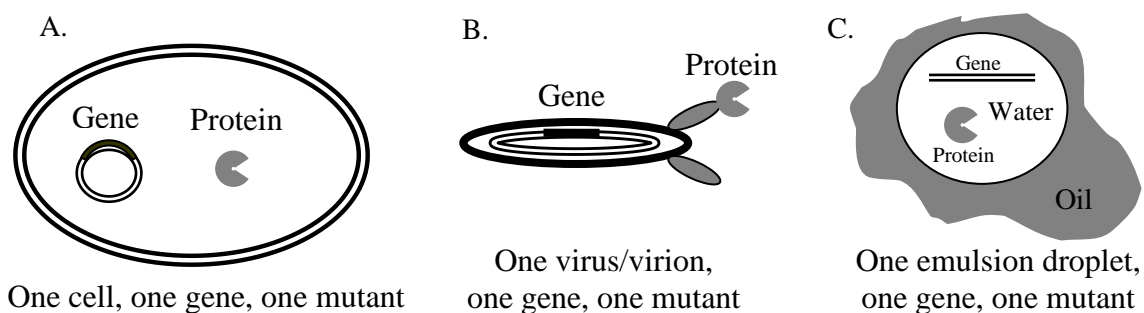


Figure A2.1 Three Basic Methods to Link Genotype and Phenotype.

Directed evolution can be reduced to three important factors: gene library creation, genotype-phenotype connection, and selection. Protein libraries are generated at the genetic level. The process typically starts by subjecting a gene to error prone PCR, which can be accomplished using normal or specially evolved polymerases under certain conditions [10]. The error rate can be tuned by altering the PCR conditions although most directed evolution approaches limit the mutation rate to only two or three mutations at the protein level [11]. Multiple rounds of selection are often performed on this library. Once the error prone PCR library stops yielding proteins with improved properties, the selected members are then recombined using a technique such as DNA shuffling [12]. Recombination techniques typically consist of cutting a gene library into predetermined fragments and then ligating the fragments back together [12, 13]. The act of ligating concentrates beneficial mutations onto a

single gene. Many other strategies build off this basic theme but will not be discussed here [14].

The second feature of directed evolution is gene (genotype) and protein (phenotype) connection [1]. One-to-one correlation of the genotype and phenotype is very important when selecting from a large library since proteins are not amplifiable. To ensure that a selected protein can be amplified, that protein must be associated with its easily amplifiable gene. The two main methods of genotype-phenotype connection are *in vivo* and *in vitro* [8]. *In vivo* methods rely on the cell to link the gene and protein, whereas *in vitro* methods are cellular and rely on a virus,[15] virion,[16, 17] or emulsion[18] to hold the gene and protein together.

*In vivo* methods rely on a whole cell to link one gene and one protein together. The process begins with the creation of a gene library that is then placed into an appropriate vector (eg. DNA plasmid). The vector is then transformed into an appropriate host cell (ie. *E. coli*). The vector is expressed within the cell and screening/selecting typically takes place without disrupting the cell [3].

*In vivo* methods differ widely in the ways that successful enzymes are culled from unsuccessful enzymes. Four techniques have been used to screen/select for successful enzymes *in vivo*: (1) Evolved enzyme yielding products that are colored/fluorescent [19], (2) Evolved enzyme replicating its own gene [20], (3) Evolved enzyme yielding a product linked to cell survival [21], and (4) Evolved enzyme creating a product that triggers transcription of a reporter gene [22]. The first three methods have been used successfully to evolve a number of protein classes, but all require a specialized substrate and as such none are general. *In vivo* methods require whole cells, and neither the protein being evolved nor the substrate being used to screen/select should adversely affect cellular processes. The adverse effects to cells can sometimes be used as the screen/select, but this is often not the case and these

adverse effects can provide misleading results.[23]

## Polymerases

Polymerases are an important class of enzyme that can synthesize DNA from a DNA template, RNA from a DNA template, and DNA from RNA template [24, 25].

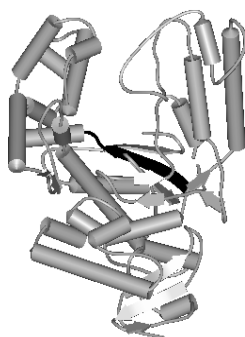


Figure A2.2 SF crystal structure with motif a highlighted in black.

These enzymes are crucial in key molecular biology experiments such as polymerase chain reaction (PCR), DNA sequencing, and genotyping [26]. This importance has prompted researchers to use directed evolution (DE) to improve thermal stability, increase substrate promiscuity, and to both increase and decrease error rate in an effort to tailor a polymerase to a specific application [27, 28]. Additionally, directed evolution and rational design of polymerases with novel functions has identified important structure function relationships in polymerases.

The mutability of the *Thermus aquaticus* (Taq) DNA Pol I active site is an example of structure function relationships that have been identified. Loeb et. al found that all 18 residues associated with Motif A (605-617) are highly mutable except for residue 610 (Figure A1.2) [29]. The flexibility of the active site led to further experimentation with Motif A to see if substitutions could allow a Taq DNA Pol I to synthesize RNA. Towards this goal, Loeb et. al made a library of 200,000 Motif A mutants and screened for activity in vivo by complementation. Twenty-three

mutants were identified that incorporated NTPs at a rate approaching  $10^3$  fold greater than that of wild-type Taq DNA Pol.

Loeb's experiments and results were an important breakthrough in the field of polymerase modification. We believe that by building on this foundation, we can learn even more about polymerases. Loeb's work will allow us to verify our proposed polymerase directed evolution method. Our system builds on that of Loeb by utilizing the increased flexibility inherent to in vitro systems. Most importantly, our method allows us to directly select for polymerases that incorporate NTPs, as opposed to Loeb's experiment which selected for complementation or incorporation of dNTPs followed by a screen for the incorporation of NTPs. This subtle difference could result in our method producing different answers than those obtained by Loeb. Also with Loeb's system, very successful mutants would be selected against since NTPs, which are available at a much higher concentration in vivo, would be incorporated into genomic DNA resulting in cell death [30]. Finally, our in vitro approach also allows us to screen more mutants than is practical for an in vivo system.

Recently, phage display has been used to modify the properties of Stoffel Fragment (SF) polymerase to allow it to incorporate non-natural monomers. SF, a truncated version of DNA pol from *T. aquaticus*, is missing the 5'→3' exonuclease domain and has a number of crystal structures available [31]. Phage display uses bacterial viruses (phage) to connect genotype to phenotype [32]. The protein of interest is attached to one of the phage coat genes, typically one of the genes encoding for a protein found on the end of the phage, pIII (Figure A1.2). Using a variety of methods the modified genome is expressed along with the other proteins to form a viable phage in a bacterial cell. Once formed, the phage displays the protein of interest on its exterior while holding the genomic information in its interior. The product of the reaction must "tag" the successful phage for evolution, which is



accomplished by Schultz and Romesberg by cleverly displaying two proteins on the surface of the phage (Figure A1.3) [6, 15]. One of the proteins serves as a “dock” for the substrate and the other protein is the protein being evolved. Once the substrate has been attached to the phage, it will be tagged and carried through the selection if the displayed enzyme can catalyze bond-formation. This phage method is spectacular, but difficult to implement and may suffer from geometric constraints if the protein active site and the substrate are not able to adopt an optimal orientation.

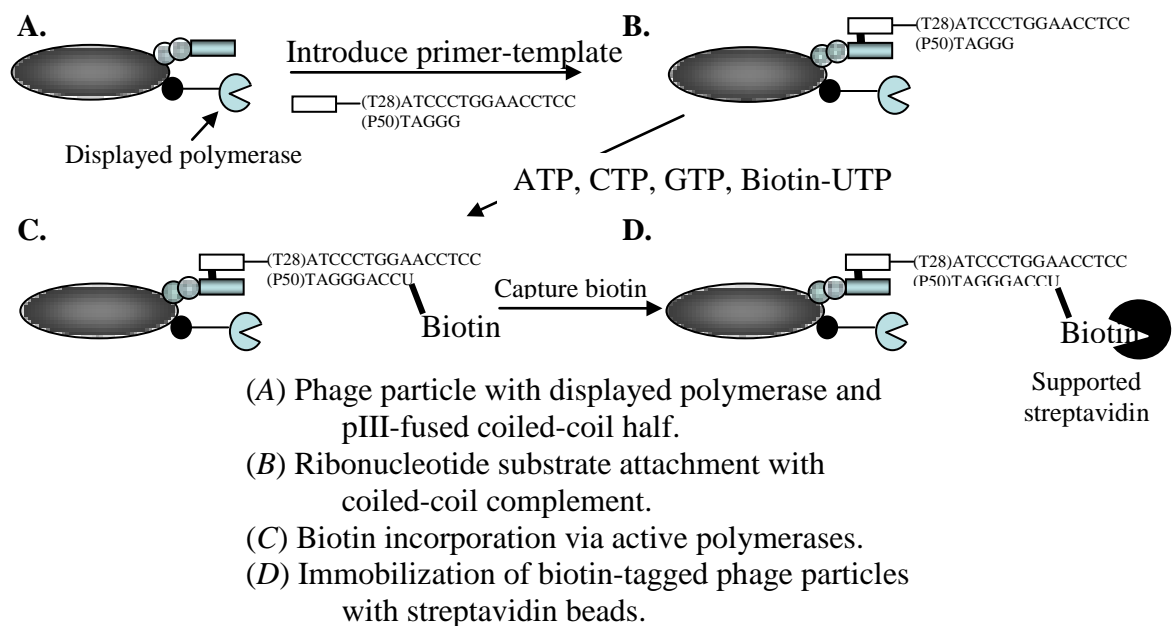


Figure A2.3. Romesberg and Schultz’s Phage Display Method for Directed Evolution of Polymerases.

This method was used to evolve SF Pol into a polymerase that could accept NTPs. After 4 rounds of evolution, several SF Pol mutants were found that had a  $10^3$  greater ribonucleotide incorporation efficiency of GTP, CTP, and ATP. Romesberg went on to use this strategy to evolve SF Pol to accept 2'-OMe NTPs. Oligonucleotides containing 2'-OMe substitutions are desirable because they are more resistant to nuclease degradation. After 4 rounds of selection from a library of  $6.4 \times$

$10^7$ , a SF Pol mutant was found that could incorporate 2'-OMe NTPs  $10^4$  times faster than the wild-type enzyme. This mutant contained I614E and E615G mutations. Alternatively, the emulsion-based approach, reported by Griffith and Tawfik, can evolve polymerases using in vitro transcription and translation reactions in a water-in-oil emulsion (Figure A2.4) [18]. As protein is expressed, the gene and protein are compartmentalized together by the boundaries of an emulsion water droplet. Ideally, one gene and multiple copies of a single mutant protein coexist per droplet. This method, termed in vitro compartmentalization (IVC), has been adapted to the

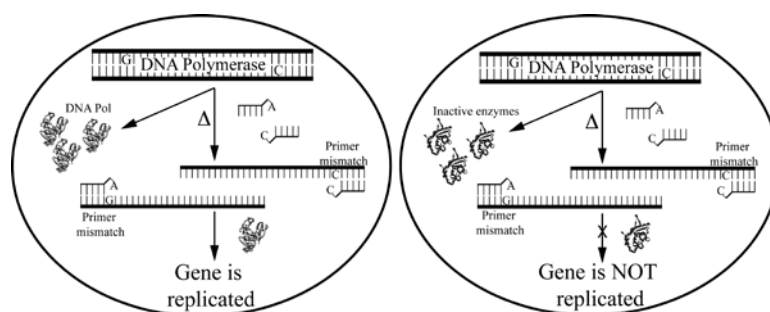


Figure A2.4. Holliger's Emulsion Approach for the Directed Evolution of a DNA polymerase.

evolution of DNA polymerases by Holliger et al. in a process called compartmentalized self replication (CSR) [20]. Holliger et al. used 3' mismatched primers to select for polymerases more tolerant of base mismatches [20]. As shown in Figure A1.3, the selection relied on the polymerase's ability to replicate its own gene using primers with a terminal 3' single mismatch.

This method yielded a polymerase with relaxed specificity allowing it to bypass blocking lesions, perform PCR with all four nucleotide phosphorothioates and incorporate base dye-labeled nucleotides at a higher frequency. Unlike phage display, Holliger's CSR approach cannot be used to select for polymerases able to replicate templates with non-natural monomers.

The overall goal of this research is to develop new methods to evolve

polymerases and proteases for specific applications. We have already achieved several key milestones that will allow us to evolve polymerases. We selected the evolution of a polymerase as our first example because a number of other directed evolution methods have been used to evolve DNA polymerases. This prior work will allow us to assess the performance of our approach through comparison to these other methods. This section delineates the success we have had so far in (1) the implementation of IVC, (2) the expression of protein within emulsions, (3) the labeling of a gene by its expressed polymerase in emulsion, (4) the design and assembly of a library, and (5) the performance of a first round of directed evolution of SF Pol (Figure A2.5)

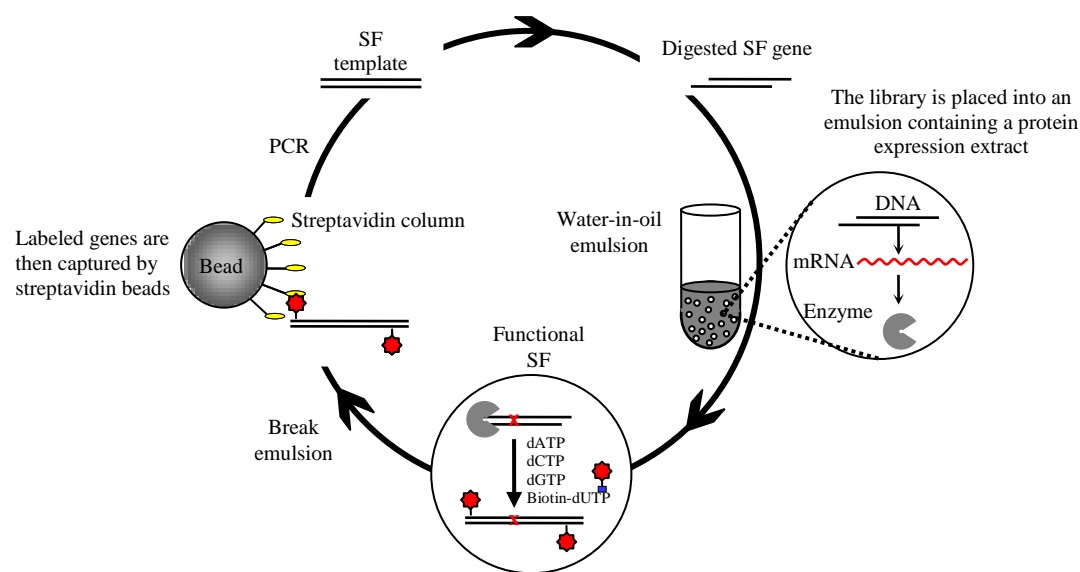


Figure A2.5. Overview of novel polymerase DE method

## ***Results and Discussion***

### **Implementation of IVC**

The in vitro expression of proteins within emulsions is a well-established area, which has been performed in several laboratories [33]. A water-in-oil emulsion is used in IVC to create a synthetic cell where a gene is expressed and confined with its

protein product in a single water droplet. Despite the wide use of IVC, stable emulsion formation can be difficult due to variations in stirring apparatus and surfactant purity [34]. Before importing IVC, we investigated several emulsion formulations to determine which provided a stable emulsion capable of synthesizing protein.

A water-in-oil emulsion consists of small water droplets (1-50 $\mu$ m in diameter) suspended in an oil phase. The emulsion is created by mixing surfactant, oil, and a small amount of water together. We examined several formulations from the literature [18, 35-38] and found that a mixture of 4.5% w/w Span 80 and 0.5% w/w Triton X-100 in mineral oil gave stable emulsions. The organic phase is prepared by dissolving Span 80 and Triton X-100 in mineral oil and cooling on ice. The aqueous phase consists of the gene to be expressed and EcoPro, an in vitro T7 RNA polymerase transcription and an *E. coli* translation extract (Novagen). The emulsion is formed by mixing the oil phase with a homogenizer at 6,500 rpm and slowly adding the aqueous phase. This method produced an emulsion stable for greater than 4 hours at 37°C.

Not only did this formulation provide a stable emulsion, it formed droplets of an appropriate size for protein synthesis. If the aqueous droplets are too large, several genes will be present in each reverse micelle leading to scrambling of genotype and phenotype. If the droplets are too small, protein synthesis could be limited by the absence of components. Statistical analysis of our emulsion from optical microscope images (not shown) revealed that our average droplet diameter was  $2\pm 1$   $\mu$ m. This size range is comparable to those found in the literature for IVC systems [18, 37].

### **Expression of proteins in emulsion**

To verify that our emulsion formulation was capable of supporting protein expression, we charged a 50 $\mu$ L EcoPro transcription and translation reaction with 1 $\mu$ g of DNA template encoding S•Tag labeled DNA polymerase I from *E. coli* and another reaction

with 100ng of template, a negative control containing no template DNA was also assembled. Emulsions were made as previously described and incubated at 37°C for 2 hours. The emulsion was then broken by adding 200  $\mu$ L of ice cold TE buffer and 1 mL of water saturated ether. After the phases were separated, the aqueous phase was washed several times with ether. A modified S•Tag assay [39] (Novagen), which is based on the assembly of a ribonuclease from the supplied S•Tag grade protein and the

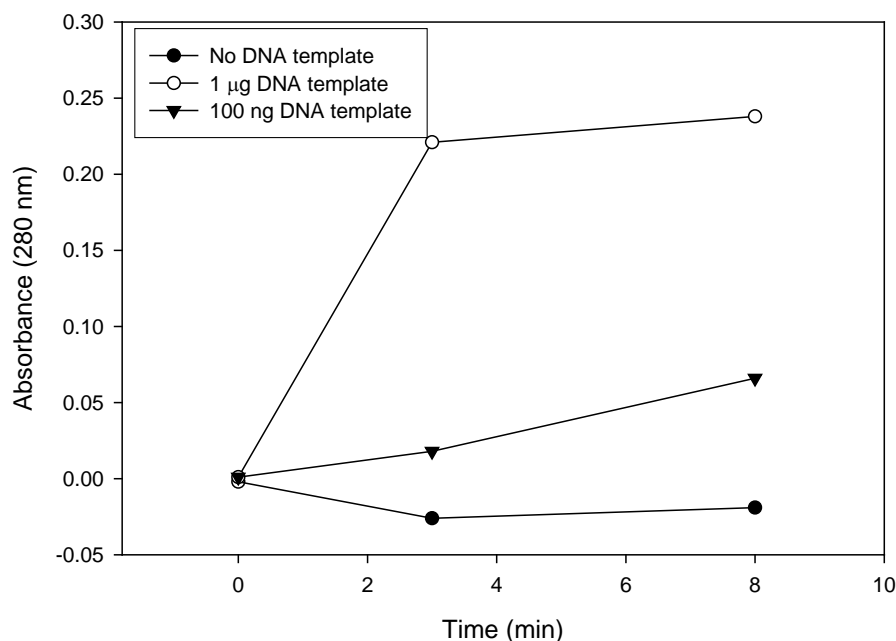


Figure A2.6 The production of protein within the water-in-oil-emulsion.

Using the modified S•Tag assay, CMP monomer production from the reconstituted ribonuclease S protein was used to indicate the presence or absence of protein synthesized. The DNA template used was plasmid, supercoiled DNA encoding a protein with the S•Tag at the C-terminus.

The graph shows a representative experiment.

S•Tag from the expressed protein, was used to assess whether protein was synthesized in the emulsions. The assay reaction containing ribonuclease substrate (polyC), S•Tag

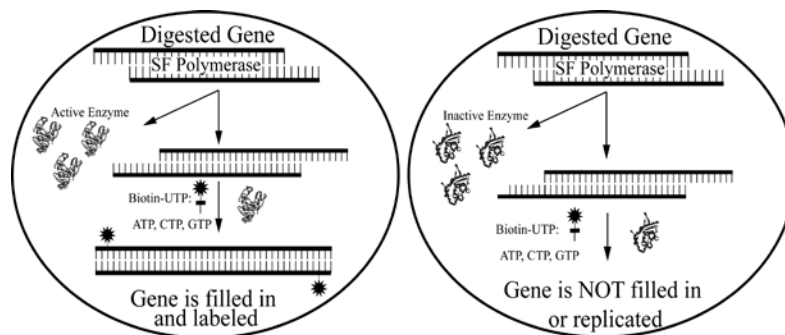


Figure A2.7. The Successful Approach SF Polymerase Selection.

grade protein, sterile water, ribonuclease buffer, and aqueous phase from the emulsion was assembled for each of the three reactions. A zero time point was removed from the assay reaction and quenched with 25% ice-cold trichloroacetic acid (TCA). Assay reactions were then incubated at 37 °C and aliquots were taken at 3 and 8 minutes and quenched. The quenched aliquots were incubated on ice and spun in a cold microcentrifuge to precipitate and pellet any remaining polyC ribonuclease substrate. The absorbance of the supernatant was measured at 280nm and used to determine how much CMP was liberated after reaction with the assembled ribonuclease. As shown in Figure A1.5, the protein synthesized in emulsion assembled with the supplied S•Tag grade protein, digested the polyC substrate, releasing CMP and caused an increase in absorbance at 280nm over time. These results indicated that our emulsion system was capable of synthesizing protein.

### **Labeling of a gene by its expressed polymerase in emulsion**

Recently, Romesberg identified a SF mutant, I614EE615G, using directed evolution and phage display, which is capable of making short strands of RNA. To determine whether a polymerase could be expressed in an active form in our emulsion system, we decided to make the SFI614EE615G mutant and try to recover its gene from our system (Figure A1.5 and A1.6).

The SF DNA Pol gene was amplified from pWB254b (ATCC#69244) by PCR utilizing primers encoding Eco RI and Hind III restriction sites. The gene was then

cloned into *E. coli* expression vector pET-29c(+) (Novagen) and confirmed with DNA sequencing. The gene was then mutated using PAGE purified oligonucleotide primers encoding the mutations I614EE615G and the QuickChange Mutagenesis kit (Stratagene). The mutation was confirmed with DNA sequencing.

To assess the activity of our constructed SF DNA Pol mutant, we attempted to recover its gene from our system. First, the SFI614EE615G gene was amplified by PCR from its vector using specially modified primers (Table A1.1). The forward primer of this set was based on the commercially available pETUP primer (Novagen) which allows for the appropriate amount of space between the end of the gene and the T7 promoter. The reverse primer was based on the T7 terminator primer (Novagen). To these primer sequences (suggested by the manufacturer), we added regions to the 5' end analogous to the sequences used in ligation independent cloning. These sequences began with a T on the 3' end and continued with 20 bp utilizing only A, C, and G.

Table A2.1. Sequences of the specialized primers that allow for partial digestion of the SF gene.

| <b>Primer</b>     | <b>Sequence (5'-3')</b>                                  |
|-------------------|--|
| Forward<br>Primer | AAC GAA GAC CAC GAC AAG ATA TGC GTC CGG CGT<br>AGA       |
| Reverse<br>Primer | AAC GAA GAG GAG AAG ACC AGT GCT AGT TAT TGC<br>TCA GCG G |

The gene is then digested in the 3'-5' direction with T4 DNA polymerase in the presence of dATP which stops the digestion at the first adenine encountered. The digested SFI614EE615G gene was then purified and used to program an in vitro transcription and translation reaction. One microgram of digested DNA was added to a 50µL reaction containing 35 µL EcoPro extract, 2µL 5 mM methionine, and 5 µL 10

mM biotin-16-UTP (Roche Applied Science). The in vitro transcription and translation reaction was then added to an oil and surfactant mixture to make an

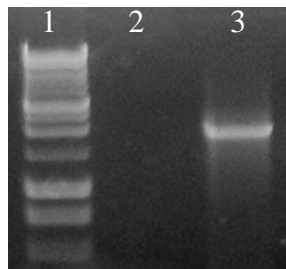


Figure A2.8. Agarose gel of biotin incorporation by SF Pol in emulsion.

**Lane 1:** 1 kb DNA ladder (Promega).

**Lane 2:** Negative control. Shows with no biotin-UTP SF gene is not labeled nor recovered a streptavidin resin.

**Lane 3:** SFI614EE615A. A band shows SF Pol incorporates biotin into its gene by polymerizing RNA, allowing for amplification from the streptavidin resin.

emulsion.

After incubation for 2 hours at 37°C, the reaction was stopped and emulsion broken by the addition of 200  $\mu$ L of ice cold TE buffer and 1 mL ice cold water saturated ether. The broken emulsion was centrifuged for 5 minutes to separate the aqueous and organic phases and the aqueous phase was washed twice with 1 mL ice cold water saturated ether. The reaction was then purified to remove unincorporated biotin-16-UTP. The DNA was then exposed to Streptavidin Paramagnetic Particles (SA-PMPs, Promega) and incubated for 30 minutes at room temperature. The SA-PMPs were washed three times with 300  $\mu$ L sterile water and resuspended in 10  $\mu$ L sterile water. PCR is then used to amplify any captured, biotinylated genes. As shown in Figure 5, PCR product was obtained from our reaction indicating that the expressed polymerase was active. The negative control, which did not contain biotin-16-UTP, showed no PCR product since it was unable to label its own gene and survive the streptavidin column. Therefore, our gene construct was able to express the SFI614EE615G protein in an active form that labeled its own gene by the incorporation of biotin-UTP.



### Construction of a SF library for directed evolution

The first step in a directed evolution experiment is to design and make a DNA library. We chose to use saturation mutagenesis to completely randomize residues 611-615 in the Motif A of Stoffel Fragment. This region is known to interact with the incoming nucleotide triphosphate during DNA synthesis. In other literature examples of the evolution of Stoffel Fragment to accept NTPs or 2'-OMe NTPs, mutations have been found in this region. Therefore, it is an ideal place to start to change SF Pol so that it can accept ribonucleotide monomers.

Table A2. 2. Mutants of Taq DNA Pol/SF found in Motif A that allow incorporation of NTPs or 2'-OMe NTPs

| Activity                    | Mutations of most active clones  | Reference     |
|-----------------------------|--|---------------|
| NTP incorporation           | <b>I614K</b> , A608: <b>I614N</b> ,<br><b>I614N:L616I</b> ,<br>A608D: <b>E615D</b> | [30]          |
| NTP incorporation           | <b>I614T:E615G</b> ,<br>A597T: <b>E615G</b>  | <sup>11</sup> |
| 2'-OMe NTP<br>incorporation | <b>I614E:E615G</b>   | [40]          |

The SF library was made using an oligonucleotide primer that encoded NNS at residues 611-615. This allowed complete randomization to all 20 amino acids, reduction in bias due to the degeneracy of the genetic code, and a reduction in the number of possible stop codons. PCR and the randomized primer were used to

amplify approximately half of the gene. The remainder of the gene was amplified with non-degenerate primers so that there was a region of overlap consisting of approximately 20 bp. Single overlap extension PCR was then used to reassemble the gene (data not shown).

#### **First round of directed evolution of SF Pol**

The assembled library (see 4, above) was digested with T4 DNA polymerase as mentioned previously. A 50 $\mu$ L in vitro transcription and translation reaction was programmed with 1  $\mu$ g of digested library. After breaking the emulsion and exposure of the aqueous layer to streptavidin paramagnetic particles, PCR was used to amplify any labeled sequences. These sequences were selected because they were able to label their own DNA with biotin, indicating that they could polymerize RNA.

These sequences were amplified and blunt ligation (Perfectly Blunt Cloning Kit, Novagen) was used to clone them into the pETBLUE-2 vector. Following transformation into chemically competent cells, individual colonies were grown, their

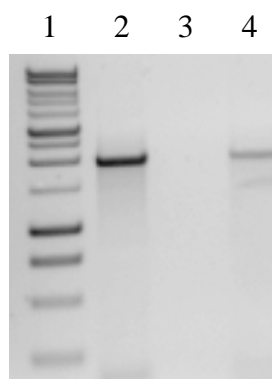


Figure A2.9. Agarose gel of successful first round of SF Pol directed evolution.

**Lane 1:** 1 kb DNA ladder.

**Lane 2:** Positive control, SFI614EE615G. A band shows mutant can label its own gene.

**Lane 3:** Negative control. No band shows that SFI614E E615G gene is not recovered in absence of biotin-16-UTP.

**Lane 4:** SF Pol library. A band shows that some mutants from the DNA library are able to label their own gene with biotin-16-UTP.

plasmids isolated and subjected to DNA sequencing. Examination of approximately 60 colonies revealed several interesting sequences. Foremost, the sequence obtained by Romesberg and known to incorporate 2'-OMe NTPs was found, indicating our method can isolate polymerase mutants capable of synthesizing non-natural oligomers. Other interesting mutants, different from those observed by Loeb and Romesberg, included sequences with a similar structural motif of an acidic residue at position 614 or 615 and a small residue at the other position [30]. Clearly, these mutants could encode polymerases that have equal or greater ability to incorporate NTPs than previously reported mutants.

Table A2.3. New mutants with the potential to polymerize NTPs.

| <b>Mutant<br/>Designation</b> | <b>611</b> | <b>612</b> | <b>613</b> | <b>614</b> | <b>615</b> |
|-------------------------------|------------|------------|------------|------------|------------|
| SLRIV256B12                   | I          | A          | M          | E          | C          |
| SLRIV256C1                    | E          | W          | R          | H          | E          |
| SLRIV256D8                    | K          | F          | G          | C          | E          |
| SLRIV256D12                   | A          | R          | E          | D          | L          |
| SLRIV264D8                    | R          | N          | Y          | E          | G          |

### ***Conclusions***

We have developed a new selection method for polymerase activity using an emulsion-based system. We have confirmed that a stable emulsion can be formed that allows for the expression of active protein. We have further proved that our system can recover genes from active polymerases. Further work including an assessment of the mutants from our first round of evolution, and subsequent rounds of evolution is definitely warranted.

## REFERENCES

1. Bommarius, A.S. and B.R. Riebel, *Biocatalysis*. 2004, betz-druck GmbH, Darmstadt: Wiley-CH.
2. Yazbeck, D.R., et al., *Challenges in the development of an efficient enzymatic process in the pharmaceutical industry*. Tetrahedron-Asymmetry, 2004. **15**(18): p. 2757-2763.
3. Arnold, F.H., *Design by directed evolution*. Accounts of Chemical Research, 1998. **31**(3): p. 125-131.
4. Reetz, M.T., *Combinatorial and evolution-based methods in the creation of enantioselective catalysts*. Angewandte Chemie-International Edition, 2001. **40**(2): p. 284-310.
5. Reetz, M.T. and K.E. Jaeger, *Superior biocatalysts by directed evolution*, in *Biocatalysis - from Discovery to Application*. 1999. p. 31-57.
6. Xia, G., et al., *Directed evolution of novel polymerase activities: Mutation of a DNA polymerase into an efficient RNA polymerase*. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(10): p. 6597-6602.
7. Kretz, K.A., et al., *Gene site saturation mutagenesis: A comprehensive mutagenesis approach*, in *Protein Engineering*. 2004. p. 3-11.
8. Lin, H.N. and V.W. Cornish, *Screening and selection methods for large-scale analysis of protein function*. Angewandte Chemie-International Edition, 2002. **41**(23): p. 4403-4425.
9. Doi, N. and H. Yanagawa, *Genotype-phenotype linkage for directed evolution and screening of combinatorial protein libraries*. Combinatorial Chemistry & High Throughput Screening, 2001. **4**(6): p. 497-509.
10. Williams, G.J., A.S. Nelson, and A. Berry, *Directed evolution of enzymes for*

- biocatalysis and the life sciences*. Cellular and Molecular Life Sciences, 2004. **61**(24): p. 3034-3046.
11. Wang, D., et al., *Estimation of the mutation rate during error-prone polymerase chain reaction*. Journal of Computational Biology, 2000. **7**(1-2): p. 143-158.
  12. Stemmer, W.P.C., *DNA Shuffling by Random Fragmentation and Reassembly - in-Vitro Recombination for Molecular Evolution*. Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(22): p. 10747-10751.
  13. Zhao, H.M., et al., *Molecular evolution by staggered extension process (StEP) in vitro recombination*. Nature Biotechnology, 1998. **16**(3): p. 258-261.
  14. Lutz, S. and W.M. Patrick, *Novel methods for directed evolution of enzymes: quality, not quantity*. Current Opinion in Biotechnology, 2004. **15**(4): p. 291-297.
  15. Pedersen, H., et al., *A method for directed evolution and functional cloning of enzymes*. Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(18): p. 10523-10528.
  16. Takahashi, T.T., R.J. Austin, and R.W. Roberts, *mRNA display: ligand discovery, interaction analysis and beyond*. Trends in Biochemical Sciences, 2003. **28**(3): p. 159-165.
  17. Yonezawa, M., et al., *DNA display of biologically active proteins for in vitro protein selection*. Journal of Biochemistry, 2004. **135**(3): p. 285-288.
  18. Tawfik, D.S. and A.D. Griffiths, *Man-made cell-like compartments for molecular evolution*. Nature Biotechnology, 1998. **16**(7): p. 652-656.
  19. Alcalde, M., E.T. Farinas, and F.H. Arnold, *Colorimetric high-throughput assay for alkene epoxidation catalyzed by cytochrome P450BM-3 variant 139-*

3. Journal of Biomolecular Screening, 2004. **9**(2): p. 141-146.
20. Ghadessy, F.J., J.L. Ong, and P. Holliger, *Directed evolution of polymerase function by compartmentalized self-replication*. Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(8): p. 4552-4557.
21. Taylor, S.V., P. Kast, and D. Hilvert, *Investigating and engineering enzymes by genetic selection*. Angewandte Chemie-International Edition, 2001. **40**(18): p. 3311-3335.
22. Lin, H.N., H.Y. Tao, and V.W. Cornish, *Directed evolution of a glycosynthase via chemical complementation*. Journal of the American Chemical Society, 2004. **126**(46): p. 15051-15059.
23. Jestin, J.L. and P.A. Kaminski, *Directed enzyme evolution and selections for catalysis based on product formation*. Journal of Biotechnology, 2004. **113**(1-3): p. 85-103.
24. Kornberg, A., *Biologic synthesis of deoxyribonucleic acid*. Science, 1960. **131**(3412): p. 1503-1508.
25. Kornberg, A., *Active center of DNA polymerase*. Science, 1969. **163**(3874): p. 1410-&.
26. Hamilton, S.C., J.W. Farchaus, and M.C. Davis, *DNA polymerases as engines for biotechnology*. Biotechniques, 2001. **31**(2): p. 370-+.
27. Goodman, M.F. and B. Tippin, *The expanding polymerase universe*. Nature Reviews Molecular Cell Biology, 2000. **1**(2): p. 101-109.
28. Goodman, M.F. and B. Tippin, *Sloppier copier DNA polymerases involved in genome repair*. Current Opinion in Genetics & Development, 2000. **10**(2): p. 162-168.
29. Shinkai, A., P.H. Patel, and L.A. Loeb, *The conserved active site motif a of*

- Escherichia coli* DNA polymerase I is highly mutable. Journal of Biological Chemistry, 2001. **276**(22): p. 18836-18842.
30. Patel, P.H. and L.A. Loeb, *Multiple amino acid substitutions allow DNA polymerases to synthesize RNA*. Journal of Biological Chemistry, 2000. **275**(51): p. 40266-40272.
  31. Lawyer, F.C., et al., *Isolation, Characterization, and Expression in Escherichia-Coli of the DNA-Polymerase Gene from Thermus-Aquaticus*. Journal of Biological Chemistry, 1989. **264**(11): p. 6427-6437.
  32. Hoess, R.H., *Protein design and phage display*. Chemical Reviews, 2001. **101**(10): p. 3205-3218.
  33. Griffiths, A.D. and D.S. Tawfik, *Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization*. Embo Journal, 2003. **22**(1): p. 24-35.
  34. Sjoblom, J., et al., *Our current understanding of water-in-crude oil emulsions. Recent characterization techniques and high pressure performance*. Advances in Colloid and Interface Science, 2003. **100**: p. 399-473.
  35. Dressman, D., et al., *Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations*. Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(15): p. 8817-8822.
  36. Ghadessy, F.J. and P. Holliger, *A novel emulsion mixture for in vitro compartmentalization of transcription and translation in the rabbit reticulocyte system*. Protein Engineering Design & Selection, 2004. **17**(3): p. 201-204.
  37. Nakano, M., et al., *Single-molecule PCR using water-in-oil emulsion*. Journal of Biotechnology, 2003. **102**(2): p. 117-124.

38. Yonezawa, M., et al., *DNA display for in vitro selection of diverse peptide libraries*. Nucleic Acids Research, 2003. **31**(19).
39. Raines, R.T., et al., *The S center dot tag fusion system for protein purification*, in *Applications of Chimeric Genes and Hybrid Proteins, Pt A*. 2000, Academic Press Inc: San Diego. p. 362-376.
40. Fa, M., et al., *Expanding the substrate repertoire of a DNA polymerase by directed evolution*. Journal of the American Chemical Society, 2004. **126**(6): p. 1748-1754.